Preslia 98: 187–207, 2026
doi: 10.23855/preslia.2026.187

Preslia

Czech Vegetation Database: an open source of vegetation-plot data for research and applications

Milan Chytrý[1,*], Irena Axmanová[1], Jan Divíšek[1,2], Klára Friesová[1], Dana Holubová[1], Salza Palpurina[3,4], Marcela Řezníčková[1], Martin Večeřa[1] & Ilona Knollová[1]

[1]Department of Botany and Zoology, Faculty of Science, Masaryk University, Kotlářská 2, CZ-61137 Brno, Czech Republic; [2]Department of Geography, Faculty of Science, Masaryk University, Kotlářská 2, CZ-61137 Brno, Czech Republic; [3]National Museum of Natural History, Bulgarian Academy of Sciences, Sofia, Bulgaria; [4]Global Biodiversity Information Facility, Europe and Central Asia Regional Support Team, Sofia, Bulgaria
[*]corresponding author: chytry@sci.muni.cz

Abstract: The Czech Vegetation Database (formerly Czech National Phytosociological Database) was founded in 1996 as a national electronic archive of vegetation-plot data. Throughout its history, it has provided data for research and applications to multiple users. In 2026, the entire database was published online as an open resource. This article describes the published open datasets. The dataset CzechVeg-Open 1 contains 117,739 vegetation-plot observations (i.e., records of species composition from vegetation plots, also called relevés) from the territory of the Czech Republic. It is provided in two formats, as Turboveg 2 files and CSV files. This dataset can be explored in the CzechVeg-DataViewer, an interactive online application developed with the shiny package in R. The dataset CzechVeg-OpenStrat 1 contains a subset of 61,282 vegetation-plot observations, which is more balanced in terms of plot density in different areas and habitat types and is better suited for national-level statistical analyses. It was prepared using an original resampling method that applies thresholds on spatial distance and species-compositional similarity between plot observations. Resampling was applied within strata defined based on principal component analysis (PCA) of 18 environmental variables. Both datasets are published and described on the database website at https://czechveg.github.io and in the Zenodo repository. In addition, the database has been published in GBIF (Global Biodiversity Information Facility), where it currently includes 2,481,138 species records. Species records from individual plot observations are grouped as Sampling events as defined by the Darwin Core standards for sharing biodiversity data. In addition to the data, the database website includes instructions for data digitization using the Turboveg 2 program and a tutorial on data handling and analysis in R with code chunks. These resources can be used to train the analysis of vegetation-plot data and to prepare open and stratified-resampled versions of other vegetation-plot databases.

Keywords: Czech Republic, open science, phytosociology, relevé, vegetation plot, vegetation survey

Received: 5 Apr 2026; Revised: 22 Jul 2026; Accepted: 26 Jul 2026; Published: 29 Sep 2026

Introduction

There is a long tradition in vegetation science of sampling species composition with abundance estimates for each species in vegetation plots (Braun-Blanquet 1964, Westhoff & van der Maarel 1978). These plot-based records were traditionally called 'relevés' and, more recently, 'vegetation-plot observations' (Dengler et al. 2011). They are used for vegetation classification, assessment of diversity patterns, study of ecological processes at the plant community level, and monitoring of vegetation and environmental changes (Schaminée et al. 2009, Jandt et al. 2022, Sperandii et al. 2022). To facilitate their use, single-time vegetation-plot observations or repeated observations from the same vegetation plot have been collected and organized in national or thematic databases (Schaminée et al. 2009, Dengler et al. 2011, Chytrý et al. 2016, Bruelheide et al. 2019, Knollová et al. 2024).

In the Czech Republic, a national database of vegetation-plot data was established in 1996 (Chytrý 1997, Chytrý & Rafajová 2003), primarily to support the development of the national vegetation classification (Chytrý 2007–2013, Chytrý & Tichý 2018). However, this database also served as one of the first examples of the use of large collections of phytosociological data to address ecological questions (Chytrý 2008) and as test data for developing analytical methods included in the JUICE program (Tichý 2002). Later, this database became part of international databases at the European scale (European Vegetation Archive; Chytrý et al. 2016) and the global scale (sPlot; Bruelheide et al. 2019).

The Czech Vegetation Database has always been understood as a resource shared by the scientific community. Many people have contributed their data, and, conversely, data from the database have been provided free of charge upon request to researchers and conservationists, both in the Czech Republic and internationally. The species occurrence data from the database were provided to the Pladias Database of the Czech Flora and Vegetation (Wild et al. 2019, Chytrý et al. 2021) and to the Species Occurrence Database of the Nature Conservation Agency of the Czech Republic (NDOP; https://portal23.nature.cz/nd).

Most recently, the Czech Vegetation Database adopted the FAIR Guiding Principles for scientific data management and stewardship (Wilkinson et al. 2016) and became an open resource available online under the CC-BY license, allowing data to be shared and adapted, as long as the source is cited and any changes are indicated. In this article, we describe the open version of the database and introduce related resources we created to facilitate the use of vegetation-plot data for science, nature conservation, ecological restoration, and education.

History of the Czech Vegetation Database

The development of the Czech Vegetation Database (formerly Czech National Phytosociological Database) was stimulated by the activities of the European Vegetation Survey Working Group of the International Association for Vegetation Science, founded in 1992. One of the aims of this group was to support national vegetation survey programs in European countries (Mucina et al. 1993) and the creation of national vegetation databases. To support database establishment, Stephan Hennekens developed the database management program Turboveg and provided it to various vegetation researchers and research groups in Europe (Schaminée & Hennekens 1995).

John Rodwell from Lancaster University, UK, proposed a project to train vegetation scientists from five central and eastern European countries, including the Czech Republic, in modern methods of vegetation survey, including the management of vegetation databases. In 1995–1998, this project, entitled "Safeguarding the Biodiversity of Eastern Europe", received funding from the British Government's Darwin Initiative (Chytrý 2026). The Czech team in this project included Milan Chytrý as a coordinator and Martin Duchoslav, Handrij Härtel, Tomáš Kučera, Kateřina Šumberová, and Lubomír Tichý as team members. This group of early-career vegetation scientists was trained at Lancaster University in 1996, in close collaboration with Stephan Hennekens. As a result of this project, the Turboveg program was prepared for implementation in the Czech Republic with the newly compiled standard species list and initial versions of the databases of bibliographic references and authors of vegetation-plot observations.

Thanks to this training and further support from John Rodwell and Stephan Hennekens, as well as cooperation on the development of the common Austrian-Czech-Slovak species list with experts from the University of Vienna (Harald Niklfeld, Walter Gutermann, and Ladislav Mucina) and the Slovak Academy of Sciences (Ivan Jarolímek and Milan Valachovič), the Czech National Phytosociological Database was established at Masaryk University in Brno in 1996 (Chytrý 1996, 1997). The database was founded by Milan Chytrý, who digitized the first few thousand vegetation-plot observations, partly by converting them, with the help of Lubomír Tichý, from an older electronic format of the FYT program, developed by Eduard Brabec at the Institute of Botany of the Czechoslovak Academy of Sciences in Průhonice in the late 1980s. The training course on vegetation databanking using Turboveg for the Czech Republic and Slovakia took place in Brno in February 1997, organized by Milan Chytrý with John Rodwell and his database assistant, Julian Dring. The course was attended by 41 participants from the Czech Republic and 8 from Slovakia. It helped to develop an organizational structure for the national vegetation database, with a central database at Masaryk University in Brno, managed by M. Chytrý, and a network of local collaborators in other important botanical institutions across the country, most notably the Institute of Botany of the Czech Academy of Sciences, Charles University in Prague, Palacký University in Olomouc, University of South Bohemia in České Budějovice, and the Czech University of Life Sciences in Prague (Chytrý 1997).

Through the collective effort of several collaborators (initially mainly David Cigánek, Jiří Danihelka, Radim Hédl, Martin Kočí, Tomáš Kučera, Zdeňka Lososová, Kateřina Šumberová, and Jana Tkáčiková), the database grew rapidly. One source of data was imports of vegetation tables created earlier in the FYT program, but most data were newly digitized from the literature and field notebooks (Fig. 1). Despite the failure of two grant applications for database funding, the database contained 6,408 vegetation-plot observations by October 1997 (Chytrý 1997). The funding was eventually secured from the Czech Science Foundation by M. Chytrý, starting from 1999, when Marie Rafajová was appointed as a full-time data manager. The grant support enabled the massive digitization of data, resulting in 55,310 vegetation-plot observations included in the database by November 2002 (Chytrý & Rafajová 2003). A considerable number of plot observations of ruderal and weed vegetation were also digitized by the team of Petr Pyšek at the Institute of Botany of the Czech Academy of Sciences for projects dealing with alien plant species.

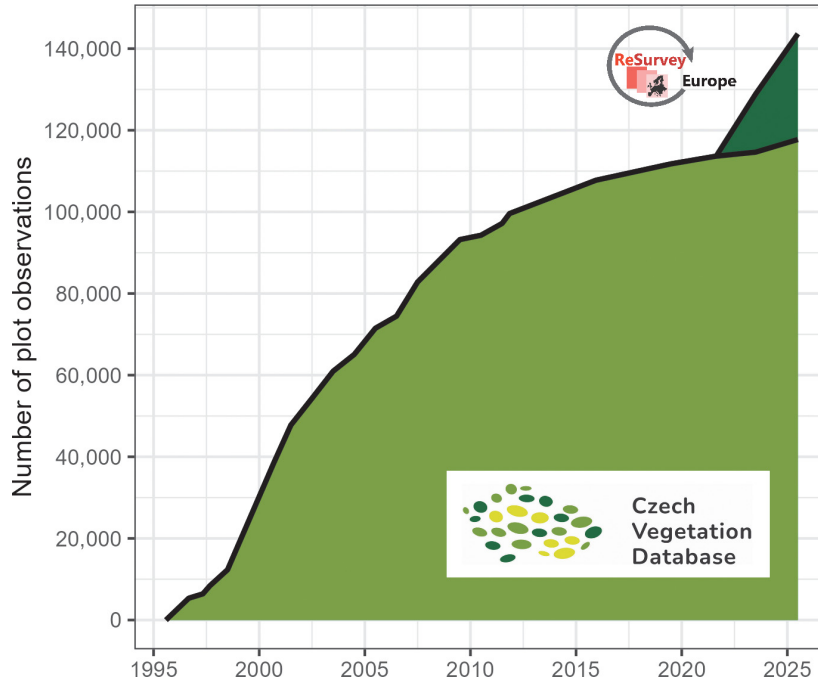

Fig. 1. The increase in the number of vegetation-plot observations in the Czech Vegetation Database (previously Czech National Phytosociological Database; lighter green) since its foundation in 1996. A steep increase began in 1999, when the first funding for the database was obtained, and lasted until 2010, when intensive data collection for the national vegetation classification project ended. Since 2022, many new plot observations from the Czech Republic have been collected for the ReSurveyEurope project (darker green); however, most of them have not been formally included in the Czech Vegetation Database. The database logo in the lower part of the graph was created in 2026.

In 2001, the database team also supported the Czech Forest Management Institute (ÚHÚL) with know-how and software to establish the Database of the Czech Forest Site Classification System as an independent database of forest vegetation plots. Forest-plot data sampled by the employees of this institute, largely from the 1950s to the 1970s, were digitized in the 2000s, resulting in a database of 39,609 forest vegetation plots and 48,977 plot observations (including repeated observations from the same plot; Zouhar 2012, https://www.givd.info/ID/EU-CZ-002).

Marie Rafajová left the team in 2002, being successively replaced by Ilona Knollová (née Kuželová) and Zdenka Preislerová (née Otýpková) (both managing the database in 2003–2005), Štěpánka Králová (2005–2008), Dana Holubová (née Michalcová; 2008–2016), and again Ilona Knollová (since 2016). These database managers coordinated data collection for the monograph Vegetation of the Czech Republic (Chytrý 2007–2013), a work supported by a series of projects from the Czech Science Foundation. Plot data were provided by project collaborators or digitized from older literature, nature conservation reports, and field notebooks. By May 2012, the database contained 99,586 vegetation-plot observations (Chytrý & Michalcová 2012).

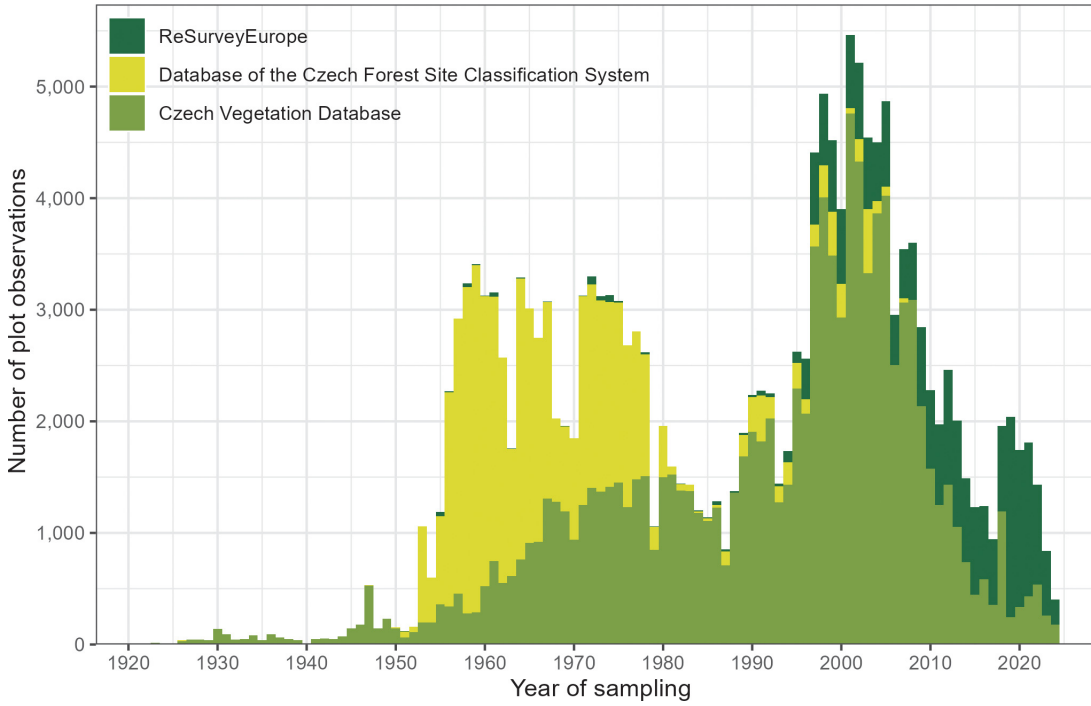

Fig. 2. Number of vegetation-plot observations in the Czech Vegetation Database by years since 1922, when the earliest plot observations are available. Plot observations from the Database of the Czech Forest Site Classification System and plot observations from the Czech Republic included in the ReSurveyEurope database are also shown for comparison.

In 2011, the Czech National Phytosociological Database was indexed in the Global Index of Vegetation-plot Databases (GIVD, https://www.givd.info, Dengler et al. 2011) with the code EU-CZ-001. In 2012, it became a founding member of the European Vegetation Archive, an integrated database of European vegetation-plot data (Chytrý et al. 2016). In 2013, it joined sPlot, the Global Vegetation Database (Bruelheide et al. 2019). Moreover, the species occurrence records from the database were imported into the Pladias Database of the Czech Flora and Vegetation, which was launched in 2018 and integrated plant species distribution records from multiple sources on a single online platform (Wild et al. 2019, Chytrý et al. 2021). These data were also imported into NDOP, a species occurrence database of the Nature Conservation Agency of the Czech Republic (https://portal23.nature.cz/nd). In return, the Nature Conservation Agency provided vegetation-plot data from Natura 2000 habitat mapping and protected area surveys.

Despite the key role of the Czech National Phytosociological Database in national and international projects of basic and applied research, the increase in the number of vegetation-plot observations has slowed after 2010, when data collection for the national vegetation classification project was completed. A new wave of data collection started after 2020, when ReSurveyEurope (Knollová et al. 2024) was launched as an international database of repeated observations from the same vegetation plots (Figs 1 and 2). However,

most ReSurveyEurope datasets have been stored separately, outside the Czech National Phytosociological Database.

Throughout its 30-year history, the database has been an important research resource. This can be documented by 300 citations of the data paper by Chytrý & Rafajová (2003) on the Web of Science and 463 citations of this and another data paper (Chytrý & Michalcová 2012) on Google Scholar by 30 June 2026. A large part of the citing publications analysed the data from the database. In addition, several dozen scientific articles used data from this database as part of broader international datasets from the European Vegetation Archive (https://euroveg.org/eva-database/publications) and the sPlot database (https://www.idiv.de/research/projects/splot/splot-publications). Species occurrence data from the database were also used in several research articles based on the Pladias Database of the Czech Flora and Vegetation.

Database opening

In 2025–2026, the authors of this article received funding from the Open Science Clusters' Action for Research and Society (OSCARS) Horizon Europe project to prepare and publish the database as a fully open online resource, serving as a good-practice example for other European vegetation-plot databases. In association with this major change, the database was renamed to the Czech Vegetation Database (CzechVeg, Česká vegetační databáze). The new name reflects that the database currently contains not only relevés that traditional phytosociologists collected for vegetation classification, but also vegetation-plot observations collected for other purposes, including ecological studies and conservation assessments. Before publishing the open version, we conducted a comprehensive quality check and updated the database, partly based on feedback from revisions of the species occurrence data from vegetation plots in the Pladias database (Wild et al. 2019), partly on various manual and semi-automatic checks.

The open version of the Czech Vegetation Database is available in two basic datasets: CzechVeg-Open includes all vegetation-plot observations from the Czech Republic stored in this database, and CzechVeg-OpenStrat is an ecologically and geographically more balanced subset of CzechVeg-Open, better suited for vegetation analyses at the national scale. In the following, we will describe these two datasets and an online tool to explore the structure of the former.

CzechVeg-Open: complete, open-access Czech Vegetation Database

CzechVeg-Open version 1 contains all 117,739 vegetation-plot observations from the Czech Republic included in the database as of 10 March 2026. Of these, 116,732 are single-time observations (relevés) from unique plots, and 1,007 are repeated observations (resurveys) from some of them. A summary of the database content in the GIVD format (https://www.givd.info) is found in Supplementary Data S1. The dataset is available in the Zenodo repository (https://doi.org/10.5281/zenodo.18289904) and can also be downloaded through the database website (https://czechveg.github.io/Data/czech_veg_open.html).

The CzechVeg-Open data are available in two formats, CSV and Turboveg 2, which contain identical sets of vegetation-plot observations. The data in the CSV format are

divided into two files: species data and header data. The species data file contains lists of species (and other taxa such as subspecies, varieties, sections, aggregates, and genera, hereafter called "species") for all plot observations, with information on the vegetation layer in which each species was recorded in each plot and its cover-abundance in this layer. Each species is also assigned to a broad plant group such as vascular plant, bryophyte, or lichen. Three different systems of species taxonomy and nomenclature are provided in the species file side by side, and users can select the one that best suits the given purpose:

(i) Turboveg species list for the Czech Republic and Slovakia, which has been used for entering species into the database. The list of vascular plant species (including some other infraspecific and infrageneric taxa) was prepared by the late Professor Harald Niklfeld and the late Dr. Walter Gutermann (University of Vienna) in 1996 based on an updated version of Liste der Gefäßpflanzen Mitteleuropas (Ehrendorfer 1973). This species list contains plant species occurring in the Czech Republic, Slovakia, and Austria. In 2015, Jiří Danihelka updated this species list to reflect additions to the Czech flora published in Danihelka et al. (2012), while keeping compatibility with the previous list. This species list is currently used by both the Czech Vegetation Database and the Central Database of Phytosociological Relevés in Slovakia (https://cbrb.sav.sk/cdf). Bryophytes and lichens included in this list follow the taxonomy and nomenclature accepted by Frey et al. (1995) and Pišút et al. (1993), respectively.

(ii) Second edition of the Key to the Flora of the Czech Republic (Kaplan et al. 2019) with a few additions as accepted in the Pladias database (https://pladias.cz/en/download/features). For bryophytes and lichens, it follows the DaLiBor database (Man et al. 2022). For taxa not available in these sources, the concepts and names are harmonized with the Catalogue of Life Extended Release taxonomy.

(iii) Expert system CzechVeg-ESy (Chytrý et al. 2020) for automatic vegetation classification according to Vegetation of the Czech Republic (Chytrý 2007–2013). This system groups some taxonomically challenging taxa into broader taxon concepts and is therefore suitable for general analyses of species composition across the database.

The header data file contains information on vegetation-plot observations other than species records and their cover-abundances. It includes details of sampling, location, vegetation structure, vegetation type, and environmental information. A detailed description of the variables included in the species and header data files is in Supplementary Data S2.

The data in the Turboveg 2 format are stored in DBF files included in a single ZIP file, which can be opened in the Turboveg 2 program (Hennekens & Schaminée 2001). For proper functioning, it is necessary to import the datasets species.zip and popup.zip, available at the Czech Vegetation Database website, into the Turboveg program. These datasets contain the Turboveg species list for the Czech Republic and Slovakia, as described above, and standardized attribute tables for some of the header data variables.

CzechVeg-DataViewer: a tool to explore the content of the Czech Vegetation Database

CzechVeg-DataViewer (https://vegscibrno.shinyapps.io/czechveg) is an interactive online application that displays plot data on the background of OpenStreetMap and produces summary graphs for selected data subsets. The subsets can be selected based on plot

geographic coordinates and their location uncertainty, plot size, year of sampling, altitude, covers of individual vegetation layers, and vegetation types according to the national vegetation classification (Chytrý 2007–2013). It is also possible to show all the plot observations that contain a selected species, a group of species, or a specified number of species from a selected species group. Header data for the selected subsets can be downloaded and used, for example, to filter data from the CzechVeg-Open for subsequent analyses. The CzechVeg-DataViewer was prepared by Klára Friesová using the shiny package in R (R Core Team 2025, Chang et al. 2026). It contains species and header data for all vegetation-plot observations that have geographical coordinates (N = 116,328 for CzechVeg-Open 1).

The CzechVeg-DataViewer provides insights into the content of the Czech Vegetation Database. The oldest plot observation in the database was sampled in 1922 (Zlatník 1928), but there are no more than 500 plot observations per year in the database until 1960 (Fig. 2). In the 1960s, research intensity increased, resulting in more than 1,000 plots sampled per year in most of the years during the 1970s and 1980s. The peak of vegetation-plot sampling in the Czech Republic occurred between 1997 and 2008, when more than 3,000 plots were sampled nearly every year. This peak was stimulated by the establishment of the database in 1996 and the data preparation for the national vegetation classification project. The initiation of national-scale Natura 2000 habitat mapping in the early 2000s (Chytrý et al. 2001, Härtel et al. 2009) also increased interest in vegetation sampling. Since 2008, the number of sampled plots per year has been declining, reflecting both a decrease in sampling intensity following the publication of the Vegetation of the Czech Republic (Chytrý 2007–2013) and the fact that more recent plot surveys have not yet been added to the database. Nevertheless, the purposes of vegetation-plot sampling have been broadened from the previous focus on describing plant community types to include various ecological studies, resurveys of historical vegetation plots to assess vegetation change, and monitoring the effects of conservation management. While 1,007 additional observations from resurveyed or monitoring plots have been included in the Czech Vegetation Database, a further 25,898 such observations were contributed to the new international database ReSurveyEurope (Knollová et al. 2024) but not included in the Czech Vegetation Database (Fig. 2).

The geographical distribution of vegetation-plot observations shows a focus on areas with well-preserved natural vegetation (Fig. 3), especially in the early periods of vegetation research, including lowland areas with a mosaic of forests and dry grasslands (České středohoří, Český kras, and South Moravia), mountain areas (Krkonoše, Hrubý Jeseník, and Šumava), and wetland areas (basins of South Bohemia). However, in the last quarter of the 20th century, sampling focused increasingly on the flysch Carpathians in eastern and southeastern Moravia, and in the 21st century, sampling covered the country more uniformly.

The most represented vegetation formations in the database, according to the national vegetation classification, are grasslands and heathlands, forests, ruderal and weed vegetation, and wetlands (Fig. 4). The most represented phytosociological classes and alliances are listed in Table 1.

The most common plot size in the database is 16 m^2 (23,970 plot observations), followed by 25 m^2 (14,980), 100 m^2 (6,541), and 400 m^2 (6,300). The first two sizes are used for sampling non-forest vegetation; 100-m^2 plots are mostly used for sampling forests,

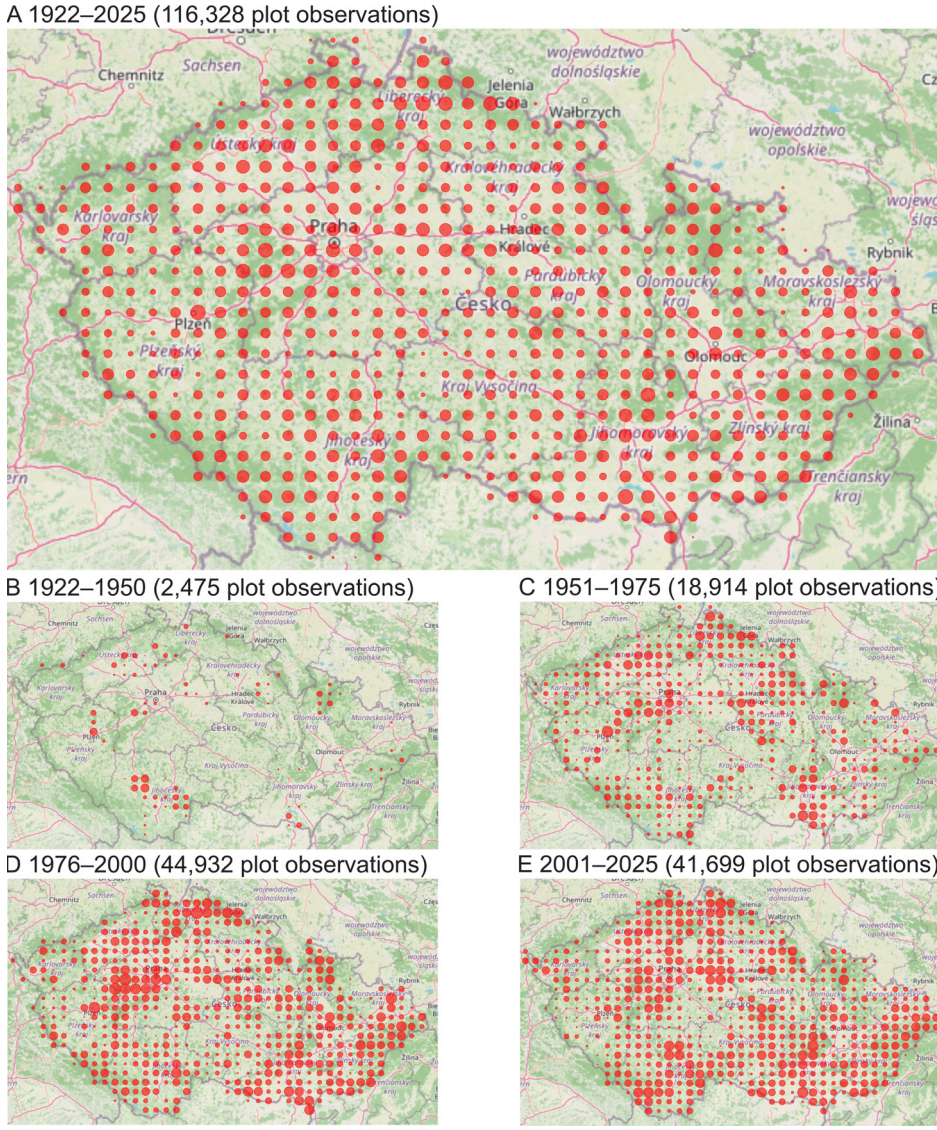

Fig. 3. The distribution of vegetation-plot observations in the Czech Vegetation Database for all plots with geographical coordinates (A) and subsets of plot observations for different time periods (B–E). The A map includes plots with missing sampling year information, whereas the B–E maps include only dated plot observations; therefore, the plot count shown in A exceeds the sum of the plot counts shown in B–E. Point sizes are proportional to the number of plot observations in KFME grid cells of 6 minutes of latitude × 10 minutes of longitude, i.e., approximately 11.1 × 12.0 km (https://pladias.cz/en/download/phytogeography).

but also used for scrub and other vegetation types, while 400-m^2 plots are nearly exclusively used for sampling forests.

Vegetation plots in the database were sampled by 737 authors and documented in 2,643 references (publications in scientific journals and books or manuscripts such as

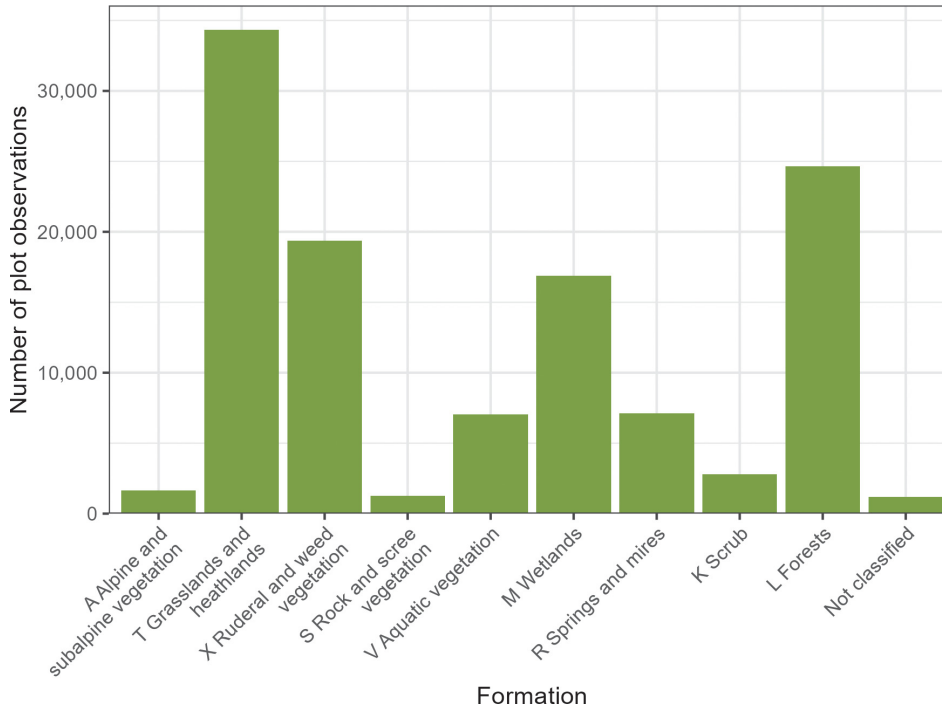

Fig. 4. Number of vegetation-plot observations by vegetation formations according to the national vegetation classification (Chytrý 2007–2013) in the Czech Vegetation Database. The classification to formations is based on the expert system CzechVeg-ESy and supplemented by subjective classification for plot observations not classified by the expert system (fields ESY_K_CODE and ESY_K_NAME; see Supplementary Data S2).

nature conservation survey reports or students' theses). The following authors sampled the highest numbers of plot observations: Jaroslav Rydlo (9,258), Richard Višňák (5,315), Emilie Balátová-Tuláčková (2,453), Kateřina Šumberová (2,204), Denisa Blažková (2,146), Jan Douda (2,100), Jiří Vicherek (1,732), Jiří Kolbek (1,692), Karel Boublík (1,519), Zdenka Preislerová (1,508), Jaromír Sofron (1,440), Jiří Sádlo (1,368), Karel Kopecký (1,344), Antonín Pyšek (1,293), Milan Chytrý (1,287), Miloslav Toman (1,246), František Grüll (1,203), Jaromír Klika (1,201), Slavomil Hejný (1,153), and Zdenka Neuhäuslová (1,141).

CzechVeg-OpenStrat: a stratified subset of the Czech Vegetation Database

As the Czech Vegetation Database is a collection of vegetation-plot data originally sampled for various purposes and using different sampling strategies, it is a highly unbalanced sample of vegetation diversity across the country. For analyses at the country scale, it is necessary to reduce biases introduced by disproportionately intensive sampling of specific locations or habitats using stratified resampling (Knollová et al. 2005, Lengyel et al. 2011, Divíšek & Chytrý 2018).

Table 1. The 20 most represented phytosociological classes (A) and alliances (B) according to the national vegetation classification (Chytrý 2007–2013) in the Czech Vegetation Database (CzechVeg-Open 1), measured by the number of vegetation-plot observations (n). The classification is based on the expert system CzechVeg-ESy and supplemented by subjective classification for plot observations not classified by the expert system (fields ESY_K_CODE and ESY_K_NAME; see Supplementary Data S2).

(A) Classes		n	(B) Alliances		n
TD	*Molinio-Arrhenatheretea*	18,421	TDF	*Calthion palustris*	6,870
LB	*Carpino-Fagetea*	16,834	TDA	*Arrhenatherion elatioris*	5,321
MC	*Phragmito-Magno-Caricetea*	13,647	MCA	*Phragmition australis*	4,980
TH	*Festuco-Brometea*	10,317	LBC	*Fagion sylvaticae*	4,534
XB	*Stellarietea mediae*	7,612	LBB	*Carpinion betuli*	3,523
XC	*Artemisietea vulgaris*	4,892	XCB	*Dauco carotae-Melilotion*	3,063
RB	*Scheuchzerio-Caricetea nigrae*	4,754	LBA	*Alnion incanae*	2,942
XD	*Galio-Urticetea*	4,320	MCH	*Magno-Caricion gracilis*	2,853
LF	*Vaccinio-Piceetea*	4,107	LBE	*Luzulo-Fagion sylvaticae*	2,763
VB	*Potametea*	4,025	LBF	*Tilio platyphylli-Acerion*	2,738
TE	*Calluno-Ulicetea*	2,951	THD	*Festucion valesiacae*	2,569
VA	*Lemnetea*	2,269	VBB	*Potamion*	2,366
KB	*Rhamno-Prunetea*	2,208	TDE	*Deschampsion cespitosae*	2,060
TF	*Koelerio-Corynephoretea*	2,193	XDE	*Aegopodion podagrariae*	2,045
LC	*Quercetea pubescentis*	1,719	THE	*Cirsio-Brachypodion pinnati*	1,988
MB	*Bidentetea tripartitae*	1,700	THF	*Bromion erecti*	1,894
XA	*Polygono arenastri-Poetea annuae*	1,667	LFC	*Piceion abietis*	1,838
MA	*Isoeto-Nano-Juncetea*	1,610	TDC	*Cynosurion cristati*	1,823
LD	*Quercetea robori-petraeae*	1,495	MCC	*Eleocharito-Sagittarion*	1,795
RC	*Oxycocco-Sphagnetea*	1,337	XBC	*Scleranthion annui*	1,642

To provide a more representative and less biased dataset, we prepared CzechVeg-OpenStrat version 1, which is a publicly available, geographically and environmentally stratified subset of 61,282 plot observations that were resampled from CzechVeg-Open 1 (https://doi.org/10.5281/zenodo.18289904, https://czechveg.github.io/Data/czech_veg_openstrat.html). This dataset, suitable for data analyses on the national and regional scales, is provided only in the CSV format, using the same data structure as the CSV version of CzechVeg-Open 1 described above.

The resampling was performed within spatially discrete environmental units (strata), which preserved the full range of environmental conditions across the Czech Republic. These strata were defined using principal component analysis (PCA; *prcomp* function, stats package, R Core Team 2025) of 18 variables describing climate (Brun et al. 2022), topography (Gessler et al. 1995, Riley et al. 1999, State Administration of Land Surveying and Cadastre 2009–2013), geology (Czech Geological Survey 1998, 2004) and soils (Ballabio et al. 2019) in 1-km grid cells of the EEA reference grid (for details see Supplementary Table S1). These variables were selected based on a preliminary PCA of a broader set of climate variables, correlation analysis (omitting highly correlated variables), and expert judgement. For each 1-km grid cell, we calculated the area of geological substrates from vector layers and mean values of other variables from raster layers. Some pixels of soil rasters (Ballabio et al. 2019) contained no data, typically in urban agglomerations or regions with extensive water bodies. These missing values were interpolated from surrounding cells using a nearest-neighbour algorithm in ArcGIS Pro (Esri 2025). Before

computing PCA, all variables were centered and scaled to a mean of 0 and a standard deviation of 1. To define environmental strata, we used the first four principal component axes, which together explained ~60% of the total variation in the dataset, and divided each axis into equal-sized intervals of 2 standard deviations. Because axes differed in length, they contained different numbers of intervals. For each 1 km grid cell, the resulting stratum was expressed as, for example, 1F_2C_3D_4E or 1E_2D_3E_4D, where numbers refer to axes (PC1–PC4) and letters to the defined intervals. This way, we obtained 314 unique strata, which were assigned to vegetation plots based on spatial overlap. The following R packages were used for data stratification: sf (Pebesma 2018), terra (Hijmans et al. 2026), dplyr, tidyr, ggplot2 (Wickham et al. 2019, 2023), and ggrepel (Slowikowski 2026). The description of this procedure, along with the R code, is available at https://czechveg.github.io/DataProcessingTutorial/data_stratification.html.

The resampling was conducted by a method proposed by Jan Divíšek (Divíšek & Chytrý 2018), which removes vegetation-plot observations that occur spatially close to other observations with similar species composition. This method was implemented in the Resampler R package available at https://github.com/jdivisek/Resampler. The package contains the *resample_plots* function, developed specifically for vegetation-plot data, which reduces plot density in areas where many plots with compositionally similar vegetation are clustered. The function first identifies the most similar pair of plot observations within the specified distance (and within a geographical or environmental stratum, if used), removes one of these plot observations based on a selected rule, and repeats this process until no pairs of plot observations that occur within the specified distance and exceed the similarity threshold remain in the dataset. A detailed description of the workflow is available at https://czechveg.github.io/DataProcessingTutorial/data_resampling.html. We used the following settings: If two plot observations were spatially closer than 1,000 m and had a species composition similarity greater than 0.6, measured by the Simpson similarity index (Simpson 1960), only the plot observation from a more recent year was retained. If both plots were from the same year, one of them was removed at random.

The resulting CzechVeg-OpenStrat dataset (Fig. 5) represents just one of many possible stratified subsets of CzechVeg-Open. Users can create alternative stratified versions, for example, by selecting different strata from the database fields containing biogeographical and landscape classifications based on Skalický (1988), Culek (1996, 2005), Chuman & Romportl (2010), Divíšek et al. (2014), included in the CSV header data of the CzechVeg-Open dataset. Another option is to select different thresholds for plot spatial proximity and species composition similarity, as well as other choices, in the *resample_plots* function.

Czech Vegetation Database in GBIF

In addition to publishing in Zenodo and on a dedicated website, the Czech Vegetation Database has also been made publicly available through GBIF (Global Biodiversity Information Facility; https://www.gbif.org/dataset/5b514765-9a1b-4910-a497-3473572ef053) using globally accepted standards for biodiversity data and metadata. This dataset (Chytrý et al. 2026) was published via the Integrated Publishing Toolkit (IPT) of the "PlantBank – CZ Botany Consortium", created by the Institute of Botany of the Czech

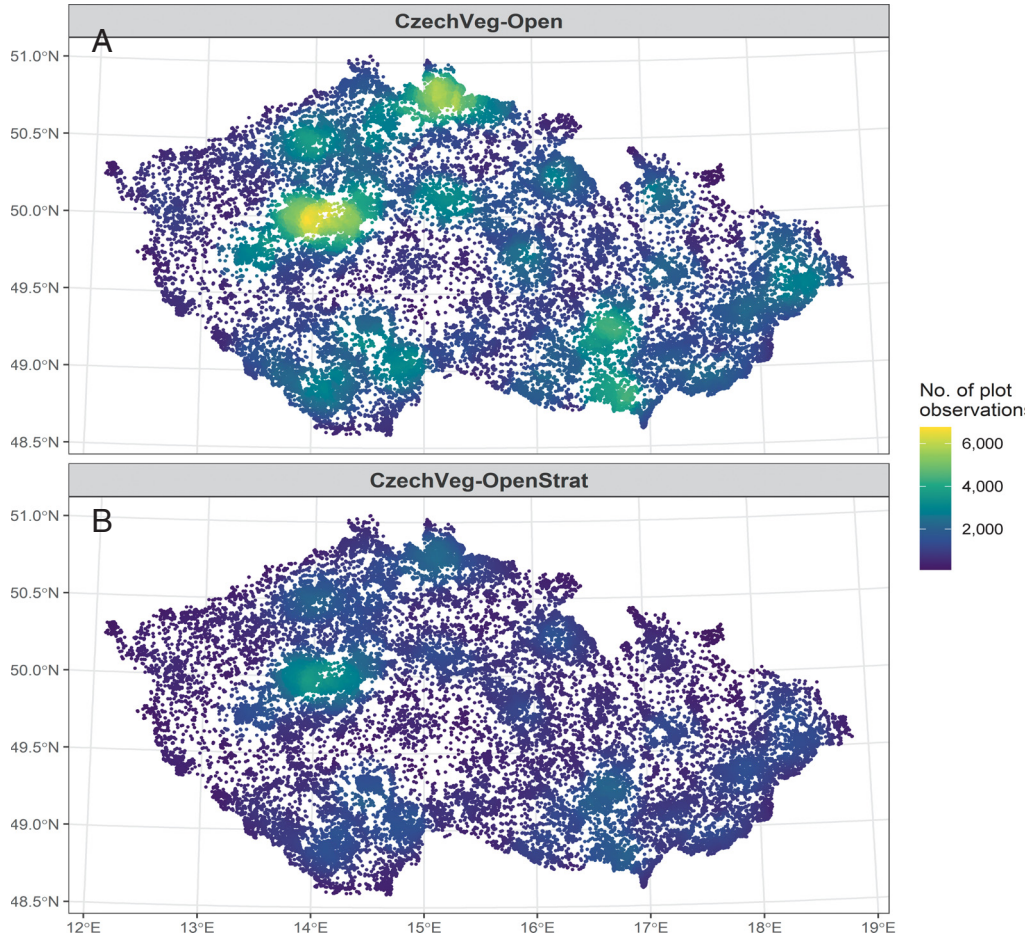

Fig. 5. Vegetation-plot observation densities in CzechVeg-Open 1 (A) and CzechVeg-OpenStrat 1 (B) datasets. In the latter dataset, the density of plot observations in overproportionally sampled areas and habitat types is considerably reduced. The density was calculated as the number of plot observations within 20 km of each plot, including those plot observations that share the same coordinates.

Academy of Sciences, Masaryk University, the University of South Bohemia, and the Czech Botanical Society in March 2026 (https://www.gbif.org/installation/57c11e59-fcd8-4053-b8d8-18084f5e02df). The PlantBank IPT archives the data as a Darwin Core Archive (DwC-A) and serves as the data repository.

The GBIF dataset of the Czech Vegetation Database is currently structured as a Sampling event dataset with three extensions: Relevé, Occurrence, and Reference. Species records (occurrences) are grouped within vegetation-plot observations, each of which corresponds to a sampling event. Version 1.15 of this dataset, published on 2 September 2026, contains 2,481,138 species occurrences.

Each vegetation-plot observation in the GBIF dataset has a unique ID (dwc:eventID according to the Darwin Core List of Terms, https://dwc.tdwg.org/list) composed of the

GIVD database code EU-CZ-001 as a prefix and the original Turboveg 2 relevé number (a unique and persistent identifier for vegetation plots in the Czech Vegetation Database). As the Turboveg 2 database does not provide identifiers for species records, we created a semi-persistent identifier as follows: EU-CZ-001-releve_nr/species_number/layer, where EU-CZ-001 is the GIVD code of the Czech Vegetation Database, releve_nr is the Turboveg 2 plot number, species_number is the Turboveg 2 species number, and layer is the indication of the vegetation layer in which the species was recorded (see Supplementary Data S2 for details of the Turboveg 2 database structure). This approach allows occurrences of the same species in different layers to be stored within a single sampling event and remains stable as long as the species identification does not change. Species names were mapped as dwc:scientificName (without authorship) and dwc:scientificNameAuthorship. Unlike the Turboveg 2 species list, the current version of the GBIF taxonomic backbone, based on the Catalogue of Life Extended Release (https://www.checklistbank.org/dataset/316165/metadata), does not support aggregates and sections as taxonomic ranks. Therefore, for the time being, names of aggregates and sections are matched to the genus or species level, and the original name is preserved as dwc:verbatimIdentification. Original species cover-abundance estimates on scales such as the Braun-Blanquet scale were converted to percentage covers across the dataset. An example workflow for converting vegetation-plot data from Turboveg 2 or other database formats into the GBIF Sampling event format is available at https://czechveg.github.io/DataProcessingTutorial/export_gbif.html.

To date, the Czech Vegetation Database is the largest botanical dataset from the Czech Republic in GBIF, comprising 74% of all GBIF vascular plant occurrences recorded in this country. It is one of the very few national vegetation-plot databases uploaded to GBIF so far, in addition to the Dutch Vegetation Database (Hennekens 2018), Ukrainian Grassland Database (Kuzemko et al. 2024) and Polish Vegetation Database (Swacha et al. 2025). However, while the Ukrainian and Polish databases are uploaded as occurrence datasets, the Dutch and Czech databases are uploaded as sampling event datasets, meaning that information on species co-occurrences and abundances in vegetation plots is preserved.

Related resources for data management and analysis

The Czech Vegetation Database is associated with other resources we developed for data management and analysis. Both data and these resources are available at https://czechveg.github.io, a website created with quarto R package (Allaire & Dervieux 2025). The source code and materials are hosted in a public GitHub repository (https://github.com/CzechVeg/CzechVeg.github.io).

Resources for data management

The Czech Vegetation Database is primarily managed using the Turboveg 2 program (Hennekens & Schaminée 2001). The page https://czechveg.github.io/Data/turboveg.html contains links and instructions for installing the program, the standard species list, attribute tables for header data fields, and detailed instructions for entering vegetation plots into the database.

Resources for data analysis

Analyses of Czech vegetation data have traditionally been conducted mainly in the JUICE program (Tichý 2002; since 2025, also with the new version called JUICE.NET, https://www.sci.muni.cz/botany/juice/NET). However, on the Czech Vegetation Database website, we provide a tutorial on basic data analyses in R (https://czechveg.github.io/DataProcessingTutorial/data_processing_tutorial.html). By doing so, we encourage users not only to adopt more flexible approaches to data analysis but also to publish code alongside each analysis, thereby improving research transparency and reproducibility in line with the principles of open science (Sperandii et al. 2024).

The tutorial shows the basic steps of data processing and visualization and provides original R code for each step, using an example dataset from the Czech Vegetation Database (65 forest understory plot observations from Axmanová et al. 2012). It explains the process of importing Turboveg files into R, handling species nomenclature, merging cover values within species across layers and across species, summarizing data (e.g. calculating plot species richness), calculating community-weighted means, working with indicator values, conversions between long and wide formats, and other commonly used procedures. This training dataset, along with all crosswalk tables, additional environmental variables, plant traits, and indicator values, is available for download directly from GitHub as a ZIP file or an RStudio project. Users are encouraged to build their own script using the copyable R code chunks, extensive commentaries, and output previews. More examples of related data analyses are available in the study materials of the Department of Botany and Zoology at Masaryk University: https://botzooldataanalysis.github.io/DataManipulationVisualisation/data_manipulation_visualisation.html. The tutorial further provides a complete description of the data stratification and resampling used to prepare the CzechVeg-OpenStrat dataset.

Conclusions

Although data from the Czech Vegetation Database have always been provided to multiple national and international users, the full opening and online publication of the database on Zenodo and GBIF have significantly increased data accessibility. We believe this will stimulate the use of these data in new areas of research and applications, including global change studies and ecological restoration. To facilitate these activities, we developed a data tutorial that demonstrates how large vegetation-plot data can be analysed in different ways.

Our database management team continues to add new vegetation plots and to check, correct, and standardize the data already included. We welcome data contributions from both individuals and institutions. In the future, we plan to publish updated, numbered versions of the database in the Zenodo repository, under the same general DOI as the current version 1 (https://doi.org/10.5281/zenodo.18289904).

As open data and other resources are crucial for scientific progress, we also encourage the managers of other vegetation-plot databases to adopt the principles of open science. The open version of the Czech Vegetation Database can serve as an example, and the published workflows and code can facilitate and accelerate the opening of other vegetation databases.

Supplementary materials

Data S1. Czech Vegetation Database overview in the standard format of the Global Index of Vegetation-plot Databases (GIVD).

Data S2. Czech Vegetation Database data description.

Table S1. Geological, climatic, topographic, and soil variables used to define environmental strata and create the CzechVeg-OpenStrat dataset.

Supplementary materials are available at https://www.preslia.cz

Acknowledgements

Credits go to hundreds of Czech botanists and vegetation scientists who collected the data in the field or processed them into a database format and provided them to the Czech Vegetation Database. Those who contributed to the database directly are listed at https://czechveg.github.io/About/acknowledgements.html. The authors of plot records are reported in the published datasets. In addition to individuals, the Czech Vegetation Database has been supported, over its history, by various institutions, most notably the Czech Science Foundation, Masaryk University, the Institute of Botany of the Czech Academy of Sciences, and the Nature Conservation Agency of the Czech Republic. We thank Petr Novotný for technical support, Pavla Růžková for facilitating the establishment of PlantBank, and Petr Petřík and Karel Prach for helpful comments on a previous manuscript version. Open publishing of the Czech Vegetation Database was supported through the Open Science Clusters' Action for Research and Society (OSCARS) Horizon Europe project under grant agreement No. 101129751, subproject "EVA-FAIR – Implementing FAIR principles in the European Vegetation Archive data".

References

Allaire J. J. & Dervieux C. (2025) quarto: R Interface to 'Quarto' Markdown Publishing System. Version 1.5.1 [Computer software]. – URL: https://doi.org/10.32614/CRAN.package.quarto.

Axmanová I., Chytrý M., Zelený D., Li C.-F., Vymazalová M., Danihelka J., Horsák M., Kočí M., Kubešová S., Lososová Z., Otýpková Z., Tichý L., Martynenko V. B., Baisheva E. Z., Schuster B. & Diekmann M. (2012) The species richness-productivity relationship in the herb layer of European deciduous forests. – Global Ecology and Biogeography 21: 657–667.

Ballabio C., Lugato E., Fernández-Ugalde O., Orgiazzi A., Jones A., Borrelli P., Montanarella L. & Panagos P. (2019) Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. – Geoderma 355: 113912.

Braun-Blanquet J. (1964) Pflanzensoziologie. Grundzüge der Vegetationskunde. – Springer, Wien.

Bruelheide H., Dengler J., Jiménez-Alfaro B., Purschke O., Hennekens S. M., Chytrý M., Pillar V. D., Jansen F., Kattge J., Sandel B., Aubin I., Biurrun I., Field R., Haider S., Jandt U., Lenoir J., Peet R. K., Peyre G., Sabatini F. M., Schmidt M., Schrodt F., Winter M., Aćić S., Agrillo E., Alvarez M., Ambarlı D., Angelini P., Apostolova I., Arfin Khan M. A. S., Arnst E., Attorre F., Baraloto C., Beckmann M., Berg C., Bergeron Y., Bergmeier E., Bjorkman A. D., Bondareva V., Borchardt P., Botta-Dukát Z., Boyle B., Breen A., Brisse H., Byun C., Cabido M. R., Casella L., Cayuela L., Černý T., Chepinoga V., Csiky J., Curran M., Ćušterevska R., Dajić Stevanović Z., De Bie E., de Ruffray P., De Sanctis M., Dimopoulos P., Dressler S., Ejrnæs R., El-Sheikh M. A. E.-R. M., Enquist B., Ewald J., Fagúndez J., Finckh M., Font X., Forey E., Fotiadis G., García-Mijangos I., de Gasper A. L., Golub V., Gutierrez A. G., Hatim M. Z., He T., Higuchi P., Holubová D., Hölzel N., Homeier J., Indreica A., Işık Gürsoy D., Jansen S., Janssen J., Jedrzejek B., Jiroušek M., Jürgens N., Kącki Z., Kavgacı A., Kearsley E., Kessler M., Knollová I., Kolomiychuk V., Korolyuk A., Kozhevnikova M., Kozub Ł., Krstonošić D., Kühl H., Kühn I., Kuzemko A., Küzmič F., Landucci F., Lee M. T., Levesley A., Li C.-F., Liu H., Lopez-Gonzalez G., Lysenko T., Macanović A., Mahdavi P., Manning P., Marcenò C., Martynenko V., Mencuccini M., Minden V., Moeslund J. E., Moretti M., Müller J. V., Munzinger J., Niinemets Ü., Nobis M., Noroozi J., Nowak A., Onyshchenko V., Overbeck G. E., Ozinga W. A., Pauchard A., Pedashenko H., Peñuelas J., Pérez-Haase A., Peterka T., Petřík P., Phillips O. L., Prokhorov V., Rašomavičius V., Revermann R., Rodwell J., Ruprecht E., Rūsiņa S., Samimi C., Schaminée J. H. J., Schmiedel U., Šibík J., Šilc U., Škvorc Ž., Smyth A., Sop T., Sopotlieva D., Sparrow B., Stančić Z., Svenning J.-C., Swacha G., Tang Z., Tsiripidis I., Turtureanu P. D., Uğurlu E., Uogintas D., Valachovič M., Vanselow K. A., Vashenyak Y., Vassilev K., Vélez-Martin E., Venanzoni R., Vibrans A. C., Violle C., Virtanen R., von Wehrden H., Wagner V., Walker D. A., Wana D., Weiher E., Wesche K., Whitfeld T.,

Willner W., Wiser S., Wohlgemuth T., Yamalov S., Zizka G. & Zverev A. (2019) sPlot – a new tool for global vegetation analyses. – Journal of Vegetation Science 30: 161–186.

Brun P., Zimmermann N. E., Hari C., Pellissier L. & Karger D. N. (2022) Global climate-related predictors at kilometer resolution for the past and future. – Earth System Science Data 14: 5573–5603.

Chang W., Cheng J., Allaire J., Sievert C., Schloerke B., Aden-Buie G., Xie Y., Allen J., McPherson J., Dipert A. & Borges B. (2026) shiny: Web Application Framework for R. R package version 1.13.0. – URL: https://CRAN.R-project.org/package=shiny.

Chuman T. & Romportl D. (2010) Multivariate classification analysis of cultural landscapes: an example from the Czech Republic. – Landscape and Urban Planning 98: 200–209.

Chytrý M. (1996) Databázový systém pro projekt přehledu vegetace České republiky [Database system for the project of the vegetation survey of the Czech Republic]. – Zprávy České botanické společnosti 31: 193–200.

Chytrý M. (1997) Česká národní fytocenologická databáze: počáteční stav a perspektivy [Czech National Phytosociological Database: initial state and perspectives]. – Zprávy České botanické společnosti, Materiály 15: 27–40.

Chytrý M. (ed.) (2007–2013) Vegetace České republiky [Vegetation of the Czech Republic] 1–4. – Academia, Praha.

Chytrý M. (ed.) (2008) 17th International Workshop European Vegetation Survey. Using phytosociological data to address ecological questions. 1–5 May 2008, Masaryk University, Brno, Czech Republic. Abstracts and Excursion Guides. – Masaryk University, Brno.

Chytrý M. (2026) The history of European vegetation databases: from relevé notebooks to a continental research infrastructure. – Annali di Botanica 16a: 7–22.

Chytrý M., Danihelka J., Kaplan Z., Wild J., Holubová D., Novotný P., Řezníčková M., Rohn M., Dřevojan P., Grulich V., Klimešová J., Lepš J., Lososová Z., Pergl J., Sádlo J., Šmarda P., Štěpánková P., Tichý L., Axmanová I., Bartušková A., Blažek P., Chrtek J. Jr., Fischer F. M., Guo W.-Y., Herben T., Janovský Z., Konečná M., Kühn I., Moravcová L., Petřík P., Pierce S., Prach K., Prokešová H., Štech M., Těšitel J., Těšitelová T., Večeřa M., Zelený D. & Pyšek P. (2021) Pladias Database of the Czech Flora and Vegetation. – Preslia 93: 1–87.

Chytrý M., Hennekens S. M., Jiménez-Alfaro B., Knollová I., Dengler J., Jansen F., Landucci F., Schaminée J. H. J., Aćić S., Agrillo E., Ambarlı D., Angelini P., Apostolova I., Attorre F., Berg C., Bergmeier E., Biurrun I., Botta-Dukát Z., Brisse H., Campos J. A., Carlón L., Čarni A., Casella L., Csiky J., Ćušterevska R., Dajić Stevanović Z., Danihelka J., De Bie E., de Ruffray P., De Sanctis M., Dickoré W. B., Dimopoulos P., Dubyna D., Dziuba T., Ejrnæs R., Ermakov N., Ewald J., Fanelli G., Fernández-González F., FitzPatrick Ú., Font X., García-Mijangos I., Gavilán R. G., Golub V., Guarino R., Haveman R., Indreica A., Işik Gürsoy D., Jandt U., Janssen J. A. M., Jiroušek M., Kącki Z., Kavgacı A., Kleikamp M., Kolomiychuk V., Krstivojević Ćuk M., Krstonošić D., Kuzemko A., Lenoir J., Lysenko T., Marcenò C., Martynenko V., Michalcová D., Moeslund J. E., Onyshchenko V., Pedashenko H., Pérez-Haase A., Peterka T., Prokhorov V., Rašomavičius V., Rodríguez-Rojo M. P., Rodwell J. S., Rogova T., Ruprecht E., Rūsiņa S., Seidler G., Šibík J., Šilc U., Škvorc Ž., Sopotlieva D., Stančić Z., Svenning J.-C., Swacha G., Tsiripidis I., Turtureanu P. D., Uğurlu E., Uogintas D., Valachovič M., Vashenyak Y., Vassilev K., Venanzoni R., Virtanen R., Weekes L., Willner W., Wohlgemuth T. & Yamalov S. (2016) European Vegetation Archive (EVA): an integrated database of European vegetation plots. – Applied Vegetation Science 19: 173–180.

Chytrý M., Knollová I., Večeřa M. & Axmanová I. (2026) Czech Vegetation Database. Version 1.15. – Masaryk University, Department of Botany and Zoology, Brno. Sampling event dataset, URL: https://doi.org/10.15468/94bfqd.

Chytrý M., Kučera T. & Kočí M. (eds) (2001) Katalog biotopů České republiky [Habitat Catalogue of the Czech Republic]. – Agentura ochrany přírody a krajiny České republiky, Praha.

Chytrý M. & Michalcová D. (2012) Czech National Phytosociological Database. – Biodiversity & Ecology 4: 345–345.

Chytrý M. & Rafajová M. (2003) Czech National Phytosociological Database: basic statistics of the available vegetation-plot data. – Preslia 75: 1–15.

Chytrý M. & Tichý L. (2018) National vegetation classification of the Czech Republic: a summary of the approach. – Phytocoenologia 48: 121–131.

Chytrý M., Tichý L., Hennekens S. M., Knollová I., Janssen J. A. M., Rodwell J. S., Peterka T., Marcenò C., Landucci F., Danihelka J., Hájek M., Dengler J., Novák P., Zukal D., Jiménez-Alfaro B., Mucina L., Abdulhak S., Aćić S., Agrillo E., Attorre F., Bergmeier E., Biurrun I., Boch S., Bölöni J., Bonari G., Braslavskaya T., Bruelheide H., Campos J. A., Čarni A., Casella L., Ćuk M., Ćušterevska R., De Bie E.,

Delbosc P., Demina O., Didukh Y., Dítě D., Dziuba T., Ewald J., Gavilán R. G., Gégout J.-C., Giusso del Galdo G. P., Golub V., Goncharova N., Goral F., Graf U., Indreica A., Isermann M., Jandt U., Jansen F., Jansen J., Jašková A., Jiroušek M., Kącki Z., Kalníková V., Kavgacı A., Khanina L., Korolyuk A. Yu., Kozhevnikova M., Kuzemko A., Küzmič F., Kuznetsov O. L., Laiviņš M., Lavrinenko I., Lavrinenko O., Lebedeva M., Lososová Z., Lysenko T., Maciejewski L., Mardari C., Marinšek A., Napreenko M. G., Onyshchenko V., Pérez-Haase A., Pielech R., Prokhorov V., Rašomavičius V., Rodríguez Rojo M. P., Rūsiņa S., Schrautzer J., Šibík J., Šilc U., Škvorc Ž., Smagin V. A., Stančić Z., Stanisci A., Tikhonova E., Tonteri T., Uogintas D., Valachovič M., Vassilev K., Vynokurov D., Willner W., Yamalov S., Evans D., Palitzsch Lund M., Spyropoulou R., Tryfon E. & Schaminée J. H. J. (2020) EUNIS Habitat Classification: expert system, characteristic species combinations and distribution maps of European habitats. – Applied Vegetation Science 23: 648–675.

Culek M. (ed.) (1996) Biogeografické členění České republiky [Biogeographical division of the Czech Republic]. – Enigma, Praha.

Culek M. (ed.) (2005) Biogeografické členění České republiky II. díl [Biogeographical division of the Czech Republic. Part II]. – Agentura ochrany přírody a krajiny České republiky, Praha.

Czech Geological Survey (1998) Geological map of the Czech Republic 1:500,000. – Czech Geological Survey, Praha.

Czech Geological Survey (2004) Geological map of the Czech Republic 1:50,000. – Czech Geological Survey, Praha.

Danihelka J., Chrtek J. Jr. & Kaplan Z. (2012) Checklist of vascular plants of the Czech Republic. – Preslia 84: 647–811.

Dengler J., Jansen F., Glöckler F., Peet R. K., De Cáceres M., Chytrý M., Ewald J., Oldeland J., Lopez-Gonzalez G., Finckh M., Mucina L., Rodwell J. S., Schaminée J. H. J. & Spencer N. (2011) The Global Index of Vegetation-Plot Databases (GIVD): a new resource for vegetation science. – Journal of Vegetation Science 22: 582–597.

Divíšek J. & Chytrý M. (2018) High-resolution and large-extent mapping of plant species richness using vegetation-plot databases. – Ecological Indicators 89: 840–851.

Divíšek J., Chytrý M., Grulich V. & Poláková L. (2014) Landscape classification of the Czech Republic based on the distribution of natural habitats. – Preslia 86: 209–231.

Ehrendorfer F. (ed.) (1973) Liste der Gefäßpflanzen Mitteleuropas. Ed. 2. – Gustav Fischer, Stuttgart.

Esri (2025) ArcGIS Pro. Version 3.6. – Environmental Systems Research Institute, Redlands.

Frey W., Frahm J.-P., Fischer E. & Lobin W. (1995) Die Moos- und Farnpflanzen Europas. – Gustav Fischer, Stuttgart.

Gessler P. E., Moore I. D., McKenzie N. J. & Ryan P. J. (1995) Soil-landscape modelling and spatial prediction of soil attributes. – International Journal of Geographical Information Systems 9: 421–432.

Härtel H., Lončáková J. & Hošek M. (eds) (2009) Mapování biotopů v České republice. Východiska, výsledky, perspektivy [Habitat mapping in the Czech Republic. Background, results, perspectives]. – Agentura ochrany přírody a krajiny České republiky, Praha.

Hennekens S. (2018) Dutch Vegetation Database. – Wageningen Environmental Research. Sampling event dataset, URL: https://doi.org/10.15468/ksqxep.

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.

Hijmans R., Brown A., Barbosa A. M., Cordano E. & Dyba K. (2026) terra: Spatial Data Analysis. Version 1.9-11. – URL: https://doi.org/10.32614/CRAN.package.terra.

Jandt U., Bruelheide H., Jansen F., Bonn A., Grescho V., Klenke R. A., Sabatini F. M., Bernhardt-Römermann M., Blüml V., Dengler J., Diekmann M., Doerfler I., Döring U., Dullinger S., Haider S., Heinken T., Horchler P., Kuhn G., Lindner M., Metze K., Müller N., Naaf T., Peppler-Lisbach C., Poschlod P., Roscher C., Rosenthal G., Rumpf S. B., Schmidt W., Schrautzer J., Schwabe A., Schwartze P., Sperle T., Stanik N., Storm C., Voigt W., Wegener U., Wesche K., Wittig B. & Wulf M. (2022) More losses than gains during one century of plant biodiversity change in Germany. – Nature 611: 512–518.

Kaplan Z., Danihelka J., Chrtek J. Jr., Kirschner J., Kubát K., Štech M. & Štěpánek J. (eds) (2019) Klíč ke květeně České republiky [Key to the flora of the Czech Republic]. Ed. 2. – Academia, Praha.

Knollová I., Chytrý M., Bruelheide H., Dullinger S., Jandt U., Bernhardt-Römermann M., Biurrun I., de Bello F., Glaser M., Hennekens S., Jansen F., Jiménez-Alfaro B., Kadaš D., Kaplan E., Klinkovská K., Lenzner B., Pauli H., Sperandii M. G., Verheyen K., Winkler M., Abdaladze O., Aćić S., Acosta A. T. R., Alignier A., Andrews C., Arlettaz R., Attorre F., Axmanová I., Babbi M., Baeten L., Baran J., Barni E., Benito-Alonso J.-L., Berg C., Bergamini A., Berki I., Boch S., Bock B., Bode F., Bonari G., Boublík K., Britton A. J., Brunet J.,

Bruzzaniti V., Buholzer S., Burrascano S., Campos J. A., Carlsson B.-G., Carranza M. L., Černý T., Charmillot K., Chiarucci A., Choler P., Chytrý K., Corcket E., Csecserits A., Cutini M., Czarniecka-Wiera M., Danihelka J., de Francesco M. C., De Frenne P., Di Musciano M., De Sanctis M., Deák B., Decocq G., Dembicz I., Dengler J., Di Cecco V., Dick J., Diekmann M., Dierschke H., Dirnböck T., Doerfler I., Doležal J., Döring U., Durak T., Dwyer C., Ejrnæs R., Ermakova I., Erschbamer B., Fanelli G., Fernández-Calzado M.-R., Fickert T., Fischer A., Fischer M., Foremnik K., Frouz J., García-González R., García-Magro D., García-Mijangos I., Gavilán R. G., Germ M., Ghosn D., Gigauri K., Gizela J., Golob A., Golub V., Gómez-García D., Gowing D., Grytnes J.-A., Güler B., Gutiérrez-Girón A., Haase P., Haider S., Hájek M., Halassy M., Harásek M., Härdtle W., Heinken T., Hester A., Humbert J.-Y., Ibáñez R., Illa E., Jaroszewicz B., Jensen K., Jentsch A., Jiroušek M., Kalníková V., Kanka R., Kapfer J., Kazakis G., Kermavnar J., Kesting S., Khanina L., Kindermann E., Kotrík M., Koutecký T., Kozub Ł., Kuhn G., Kutnar L., La Montagna D., Lamprecht A., Lenoir J., Lepš J., Leuschner C., Lorite J., Madsen B., Magaña Ugarte R., Malicki M., Maliniemi T., Máliš F., Maringer A., Marrs R., Matesanz S., Metze K., Meyer S., Millett J., Mitchell R. J., Moeslund J. E., Moiseev P., Morra di Cella U., Mudrák O., Müller F., Müller N., Naaf T., Nagy L., Napoleone F., Nascimbene J., Navrátilová J., Ninot J. M., Niu Y., Normand S., Ogaya R., Onipchenko V., Orczewska A., Ortmann-Ajkai A., Pakeman R. J., Pardo I., Pätsch R., Peet R. K., Penuelas J., Peppler-Lisbach C., Pérez-Hernández J., Pérez-Haase A., Petraglia A., Petřík P., Pielech R., Piórkowski H., Pladevall-Izard E., Poschlod P., Prach K., Praleskouskaya S., Prokhorov V., Provoost S., Puşcaş M., Pustková Š., Randin C. F., Rašomavičius V., Reczyńska K., Rédei T., Řehounková K., Richner N., Risch A. C., Rixen C., Rosbakh S., Roscher C., Rosenthal G., Rossi G., Rötzer H., Roux C., Rumpf S. B., Ruprecht E., Rūsiņa S., Sanz-Zubizarreta I., Schindler M., Schmidt W., Schories D., Schrautzer J., Schubert H., Schuetz M., Schwabe A., Schwaiger H., Schwartze P., Šebesta J., Seiler H., Šilc U., Silva V., Šmilauer P., Šmilauerová M., Sperle T., Stachurska-Swakoń A., Stanik N., Stanisci A., Steffen K., Storm C., Stroh H. G., Sugorkina N., Świerkosz K., Świerszcz S., Szymura M., Teleki B., Thébaud G., Theurillat J.-P., Tichý L., Treier U. A., Turtureanu P. D., Ujházy K., Ujházyová M., Ursu T. M., Uziębło A. K., Valkó O., Van Calster H., Van Meerbeek K., Vandevoorde B., Vandvik V., Varricchione M., Vassilev K., Villar L., Virtanen R., Vittoz P., Voigt W., von Hessberg A., von Oheimb G., Wagner E., Walther G.-R., Wellstein C., Wesche K., Wilhelm M., Willner W., Wipf S., Wittig B., Wohlgemuth T., Woodcock B. A., Wulf M. & Essl F. (2024) ReSurveyEurope: a database of resurveyed vegetation plots in Europe. – Journal of Vegetation Science 35: e13235.

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.

Kuzemko A., Vynokurov D., Vasheniak I., Budzhak V., Chorney I., Chusova O., Dembicz I., Dengler J., Didukh Y., Dziuba T., Hájek M., Janišová M., Kolomiychuk V., Konaykova V., Moysiyenko I., Roleček J., Savchenko G., Shapoval V., Shyriaeva D., Škodová I. & Tokariuk A. (2024) Records of vascular plants, bryophytes and lichens from Ukrainian Grassland Database. Version 1.3. – M.G. Kholodny Institute of Botany, National Academy of Sciences of Ukraine, Kyiv. Occurrence dataset, URL: https://doi.org/10.15468/w4yz9s.

Lengyel A., Chytrý M. & Tichý L. (2011) Heterogeneity-constrained random resampling of phytosociological databases. – Journal of Vegetation Science 22: 175–183.

Man M., Malíček J., Kalčík V., Novotný P., Chobot K. & Wild J. (2022) DaLiBor: Database of Lichens and Bryophytes of the Czech Republic. – Preslia 94: 579–605.

Mucina L., Rodwell J. S., Schaminée J. H. J. & Dierschke H. (1993) European Vegetation Survey: current state of some national programmes. – Journal of Vegetation Science 4: 429–438.

Pebesma E. (2018) Simple Features for R: Standardized support for spatial vector data. – The R Journal 10: 439–446.

Pišút I., Lackovičová A. & Lisická E. (1993) Súpis lišajníkov Slovenska [A list of lichens of Slovakia]. – Biológia 48: 53–98.

R Core Team (2025) R: A language and environment for statistical computing. – R Foundation for Statistical Computing, Vienna. URL: https://www.R-project.org.

Riley S. J., DeGloria S. D. & Elliot R. (1999) A terrain ruggedness index that quantifies topographic heterogeneity. – Intermountain Journal of Sciences 5: 23–27.

Schaminée J. H. J. & Hennekens S. M. (1995) Update of the installation of Turboveg in Europe. – Annali di Botanica 53: 159–161.

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.

Simpson G. G. (1960) Notes on the measurement of faunal resemblance. – American Journal of Science 258-A: 300–311.

Skalický V. (1988) Regionálně fytogeografické členění [Regional phytogeographic division]. – In: Hejný S., Slavík B., Chrtek J., Tomšovic P. & Kovanda M. (eds), Květena České socialistické republiky [Flora of the Czech Socialist Republic] 1: 103–121, Academia, Praha.

Slowikowski K. (2026) ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.9.8. – URL: https://doi.org/10.32614/CRAN.package.ggrepel.

Sperandii M. G., Bazzichetto M., Mendieta-Leiva G., Schmidtlein S., Bott M., de Lima R. A. F., Pillar V. D., Price J. N., Wagner V. & Chytrý M. (2024) Towards more reproducibility in vegetation research. – Journal of Vegetation Science 35: e13224.

Sperandii M. G., de Bello F., Valencia E., Götzenberger L., Bazzichetto M., Galland T., E-Vojtkó A., Conti L., Adler P. B., Buckley H., Danihelka J., Day N. J., Dengler J., Eldridge D. J., Estiarte M., García-González R., Garnier E., Gómez-García D., Hallett L., Harrison S., Herben T., Ibáñez R., Jentsch A., Juergens N., Kertész M., Kimuyu D. M., Klumpp K., Le Duc M., Louault F., Marrs R. H., Ónodi G., Pakeman R. J., Pärtel M., Peco B., Peñuelas J., Rueda M., Schmidt W., Schmiedel U., Schuetz M., Skalova H., Šmilauer P., Šmilauerová M., Smit C., Song M.-H., Stock M., Val J., Vandvik V., Wesche K., Wiser S. K., Woodcock B. A., Young T. P., Yu F.-H., Wolf A. A., Zobel M. & Lepš J. (2022) LOTVS: a global collection of permanent vegetation plots. – Journal of Vegetation Science 33: e13115.

State Administration of Land Surveying and Cadastre (2009–2013) ZABAGED® - Altimetry - DMR 5G. Digital Terrain Model of the Czech Republic of the 5th generation (DMR 5G) in S-JTSK, Bpv. – Czech Office for Surveying, Mapping and Cadastre, Praha.

Swacha G., Kącki Z. & Stefaniak J. (2025) Polish Vegetation Database. Version 1.4. – University of Wrocław. Occurrence dataset, URL: https://doi.org/10.15468/fzugec.

Tichý L. (2002) JUICE, software for vegetation classification. – Journal of Vegetation Science 13: 451–453.

Westhoff V. & van der Maarel E. (1978) The Braun-Blanquet approach. – In: Whittaker R. H. (ed.), Classification of plant communities, p. 287–399, Junk, The Hague.

Wickham H., Averick M., Bryan J., Chang W., McGowan L. D., François R., Grolemund G., Hayes A., Henry L., Hester J., Kuhn M., Pedersen T. L., Miller E., Bache S. M., Müller K., Ooms J., Robinson D., Seidel D. P., Spinu V., Takahashi K., Vaughan D., Wilke C., Woo K. & Yutani H. (2019) Welcome to the Tidyverse. – Journal of Open Source Software 4: 1686.

Wickham H., Çetinkaya-Rundel M. & Grolemund G. (2023) R for Data Science. Ed. 2. – O'Reilly Media, Sebastopol.

Wild J., Kaplan Z., Danihelka J., Petřík P., Chytrý M., Novotný P., Rohn M., Šulc V., Brůna J., Chobot K., Ekrt L., Holubová D., Knollová I., Kocián P., Štech M., Štěpánek J. & Zouhar V. (2019) Plant distribution data for the Czech Republic integrated in the Pladias database. – Preslia 91: 1–24.

Wilkinson M. D., Dumontier M., Aalbersberg I. J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.-W., da Silva Santos L. B., Bourne P. E., Bouwman J., Brookes A. J., Clark T., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C. T., Finkers R., Gonzalez-Beltran A., Gray A. J. G., Groth P., Goble C., Grethe J. S., Heringa J., 't Hoen P. A. C., Hooft R., Kuhn T., Kok R., Kok J., Lusher S. J., Martone M. E., Mons A., Packer A. L., Persson B., Rocca-Serra P., Roos M., van Schaik R., Sansone S.-A., Schultes E., Sengstag T., Slater T., Strawn G., Swertz M. A., Thompson M., van der Lei J., van Mulligen E., Velterop J., Waagmeester A., Wittenburg P., Wolstencroft K., Zhao J. & Mons B. (2016) The FAIR Guiding Principles for scientific data management and stewardship. – Scientific Data 3: 160018.

Zlatník A. (1928) Études écologiques et sociologiques sur le *Sesleria coerulea* et le Seslerion calcariae en Tchécoslovaquie. – Rozpravy Královské české společnosti nauk, třída mathematicko-přírodovědecká 8: 1–115.

Zouhar V. (2012) Database of Czech Forest Classification System. – Biodiversity & Ecology 4: 346.

Česká vegetační databáze: otevřený zdroj dat z vegetačních snímků pro výzkum a aplikace

Česká vegetační databáze (dříve Česká národní fytocenologická databáze) byla založena v roce 1996 jako národní elektronický archiv vegetačních snímků. Během své historie poskytovala na požádání data mnoha uživatelům pro výzkum a aplikace. V roce 2026 byla celá databáze zveřejněna online. Datová sada CzechVeg-Open 1 obsahuje 117 739 vegetačních snímků z území České republiky. Je poskytována ve dvou formátech: jednak jako soubory Turboveg 2, jednak jako soubory CSV. Tuto datovou sadu lze prohlížet v interaktivní online aplikaci CzechVeg-DataViewer, vyvinuté pomocí knihovny shiny v jazyce R. Datová sada CzechVeg-OpenStrat 1 obsahuje podmnožinu 61 282 vegetačních snímků, která je vyváženější z hlediska hustoty snímků v různých územích a typech stanovišť, a proto je vhodnější pro statistické analýzy na národní úrovni. Byla připravena pomocí originální metody převzorkování databáze, která aplikuje prahové hodnoty na prostorovou vzdálenost a podobnost druhového složení mezi vegetačními snímky. Převzorkování bylo provedeno v rámci strat definovaných na základě analýzy hlavních komponent (PCA) 18 proměnných prostředí. Obě datové sady jsou zveřejněny v repozitáři Zenodo a dostupné na webových stránkách databáze na adrese https://czechveg.github.io. Databáze byla navíc publikována v GBIF (Global Biodiversity Information Facility), kde obsahuje 2 481 138 druhových záznamů seskupených po jednotlivých vegetačních snímcích jako tzv. sampling events. Kromě volně stažitelných datových sad obsahují webové stránky databáze také návody pro digitalizaci dat pomocí programu Turboveg a studijní materiály o zpracování a analýze dat v jazyce R s částmi kódu. Tyto zdroje jsou použitelné pro výuku analýzy vegetačních snímků a přípravu otevřených nebo stratifikovaně převzorkovaných verzí jiných vegetačních databází.

How to cite: Chytrý M., Axmanová I., Divíšek J., Friesová K., Holubová D., Palpurina S., Řezníčková M., Večeřa M. & Knollová I. (2026) Czech Vegetation Database: an open source of vegetation-plot data for research and applications. – Preslia 98: 187–207.

Preslia, a journal of the Czech Botanical Society

https://www.preslia.cz