Plant distribution data for the Czech Republic integrated in the Pladias database

Údaje o rozšíření cévnatých rostlin v České republice integrované v databázi Pladias

Jan Wild[1], Zdeněk Kaplan[1], Jiří Danihelka[1,2], Petr Petřík[1], Milan Chytrý[2], Petr Novotný[1,3], Martin Rohn[4], Václav Šulc[1], Josef Brůna[1], Karel Chobot[5], Libor Ekrt[6], Dana Holubová[2], Ilona Knollová[2], Petr Kocián[7], Milan Štech[6], Jan Štěpánek[1] & Václav Zouhar[8]

[1]*The Czech Academy of Sciences, Institute of Botany, Zámek 1, CZ-252 43 Průhonice, Czech Republic, e-mail: jan.wild@ibot.cas.cz, kaplan@ibot.cas.cz, petr.petrik@ibot.cas.cz, josef.bruna@ibot.cas.cz, vaclav.sulc@ibot.cas.cz, jan.stepanek@ibot.cas.cz;* [2]*Department of Botany and Zoology, Masaryk University, Kotlářská 2, CZ-611 37 Brno, Czech Republic, e-mail: danihel@sci.muni.cz, chytry@sci.muni.cz, danmich@sci.muni.cz, ikuzel@sci.muni.cz;* [3]*Department of Teaching and Didactics of Biology, Faculty of Science, Charles University, Viničná 7, CZ-128 00 Praha, Czech Republic, e-mail: krkabol@google.com;* [4]*Babická 2379/1a, CZ-149 00 Praha, Czech Republic, e-mail: martin.rohn@gmail.com;* [5]*Nature Conservation Agency of the Czech Republic, Kaplanova 1, CZ-140 00 Praha 4, Czech Republic, e-mail: karel.chobot@nature.cz;* [6]*Department of Botany, Faculty of Science, University of South Bohemia, Branišovská 1760, CZ-370 05 České Budějovice, Czech Republic, email: libor.ekrt@gmail.com, stech@prf.jcu.cz;* [7]*Nerudova 5, CZ-741 01 Nový Jičín, Czech Republic; e-mail: petr.kocian@kvetenacr.cz;* [8]*Forest Management Institute, Vrázova 1, CZ-616 00 Brno, Czech Republic, e-mail: zouhar.vaclav@uhul.cz*

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.

Research on the Czech flora has a long tradition and yielded a large number of records on the occurrence of plants. Several independent electronic databases were established during the last three decades in order to collect and manage these records. However, this fragmentation and the different characteristics of each database strongly limit the utilization and analyses of plant distribution data. Solving these problems was one of the aims of the Centre of Excellence PLADIAS (Plant Diversity Analysis and Synthesis, 2014–2018), which is also the source of the name of the central database of the project: Pladias – Database of the Czech Flora and Vegetation (www.pladias.cz). We developed an occurrence module as a part of the Pladias database in order to integrate species occurrence data on vascular plants in the Czech Republic for use in pure and applied research. In this paper, we present a description of the structure of this database, data handling and validation, creation of distribution maps based on critically evaluated records as well as descriptions of the original databases and explorative analyses of spatiotemporal and taxonomic coverage of the integrated occurrence data. So far we have integrated more than 13 million records of almost 5 thousand taxa (species, subspecies, varieties and hybrids), which came from five large national databases, seven regional projects and records collected within the PLADIAS project. The Pladias database is now the largest set of data on vascular plant occurrence in the Czech Republic, which is subject to continuous quality control. Analyses of this database pointed

doi: 10.23855/preslia.2019.001

to differences in spatial and taxonomic coverage of the source datasets. However, it also showed that the targeted effort of experts focused on validating existing records, as well as the collection of new data is still necessary in order to obtain reliable distribution data for individual species.

Keywords: database, distribution, ecoinformatics, flora, grid maps, plant diversity, plant occurrence record, sampling effort, vascular plants

Introduction

Current research in biology, ecology and nature conservation is inconceivable without information on the distributions of species. Data on species occurrence have been collected for different purposes and by different people ranging from amateur collation for personal use to systematic collective efforts, often in the form of grid-based mapping at local (e.g. Grulich 1997, Kolbek et al. 1999, Jongepier & Pechanec 2006), national (e.g. Haeupler 2005, Preston 2013) or even continental scales (e.g. Atlas Florae Europaeae; Jalas & Suominen 1972 et seq.). Large volumes of such data have recently been assembled in electronic databases, which support their easy sharing for a broader cooperation among scientists and also for providing the public with information.

An example of sharing species distribution databases built up for various purposes is the largest world-wide portal for species occurrence data, the Global Biodiversity Information Facility (GBIF 2015). It succeeded in collecting species distribution records from a variety of sources (244,888,893 plant records), and provided datasets for a large number of studies (3,433 publications, www.gbif.org, December 2018). Similar initiatives for vegetation-plot data include the European Vegetation Archive (EVA; Chytrý et al. 2016) and sPlot, the global vegetation-plot database (Bruelheide et al. 2019). However, even these huge databases are incomplete and have geographic or taxonomic biases (Graham et al. 2008, Franklin et al. 2016, Meyer et al. 2016). Even at national scales plant distribution data are often dispersed among many different databases and hardcopy documents, and their integration still remains a challenging task. A successful integration of numerous databases of vascular plant records at the national scale has recently been achieved in Germany, based on joint efforts of the voluntary association Netzwerk Phytodiversität Deutschland and the state nature conservancy Bundesamt für Naturschutz, resulting in a national database (https://deutschlandflora.de) with about 30 million records of about 7,000 accepted taxa (Netzwerk Phytodiversität Deutschland & Bundesamt für Naturschutz 2013). Apart from their large contribution to the understanding of patterns in the distribution of species, large projects involving the integration of databases also reveal multifaceted issues related to data quality (Graham et al. 2008, Yang et al. 2013, Beck et al. 2014). As a reflection of this experience, current trends in ecoinformatics highlight needs for standardization, effective management and validation of the databases.

Research on the Czech flora has a long tradition in this country, dating back to the late 18th century (Kaplan 2012, Krahulec 2012, Danihelka et al. 2017) and has yielded a large number of plant records. For instance, there are about 8 million herbarium sheets stored in the public collections of about 70 institutions (http://www.mzm.cz/seznam-herbarovych-sbirek-v-cr), of which about 80% may have originated from this country. In the last two decades, millions of observations of plant occurrences were recorded in field surveys organized by the state nature conservancy or in local recording schemes, and these records are mostly stored in electronic formats.

Many independent electronic databases were established during the last three decades to collect and manage plant species records, and the scope of four of them is national. An electronic database of vascular plant records from the Czech Republic (Database of the Distribution of Vascular Plants; FLDOK) was established at the Institute of Botany of the Czech (then Czechoslovak) Academy of Sciences in 1992 (Brabec 1999). The Nature Conservation Agency of the Czech Republic (NCA CR) has been building its own information system for collecting biological records at the national scale since the mid-1980s (Podhajská & Škapec 1983). In 1996, the Czech National Phytosociological Database (CNPD, Chytrý & Rafajová 2003) was established, storing data from vegetation plots, which are also an important source of records of individual species. Masaryk University in Brno acts as a coordinator of this database. A similar database of forest vegetation plots sampled by forestry researchers (The Database of Czech Forest Classification System; DCFCS) was established by the Forest Management Institute (Zouhar 2012). The growing availability of database software supported systematic data collection activities mainly at branches of the Czech Botanical Society, local museums, nature conservation institutions, universities, research institutes and among amateur botanists. Several regional grid mapping projects have been created since the 1990s (see Petřík 2006a for a review).

Thus, currently four institutions independently manage national databases of the distribution records of plant species, and several other institutions, voluntary associations and private persons (usually amateur botanists) collate data at local or regional scales (see Appendix). Each of these databases was established with a specific, often ad hoc purpose, and has its own taxonomic backbone, structure, management regime and emphasis on particular types of records. Databases built-up by voluntary associations and individuals sometimes have a narrower geographic scope or focus on particular taxonomic groups. This inevitably causes differences in spatial coverage, record accuracy, species recorded and time scale covered. Many records lack some basic attributes, such as geographic coordinates (with accurate information) or an unequivocal assignment to a particular field of the mapping grid, author of the record, reference to the source of the record (publication or a herbarium specimen), date of record or information on occurrence status. In most records, the taxonomic information is limited solely to a scientific name that is not linked to a particular taxon in a taxonomic reference work. Moreover, the majority of records have never been checked for geographic accuracy and mistakes in the identification of taxa. One of the major prerequisites for sound data usage is thus validation by experts.

There have been two main attempts to integrate Czech databases on the distribution of plants. The web portal www.florabase.cz was launched in 2009 as an output of the project Biodiversity Research Centre (LC06073), where eight Czech scientific and academic institutions shared their digitized data and presented them to the public. Feedback on the occurrence records was collected from field botanists; however, there were no resources to use this feedback for the systematic and critical validation of the data. This portal brought together about 800 registered users and provided invaluable experience in integrating different data sources.

Independently of academic institutions, the NCA CR developed a powerful system for the collection and administration of data on species occurrence (NDOP). It also aspired to integrate records at the country scale, mainly for nature conservation purposes, resulting in the largest and still growing database of occurrences of organisms including vascular plants in the Czech Republic; the records are available online (http://portal.nature.cz/nd).

In this paper, we present a new framework to integrate records on the distribution of individual species of vascular plants in the Czech Republic (previously temporarily called CzechDistrib database), which was developed as a part of the Pladias Database of the Czech Flora and Vegetation. The name is derived from an acronym of the Centre of Excellence PLADIAS (Plant Diversity Analysis and Synthesis; www.pladias.cz), funded by the Czech Science Foundation in 2014–2018. Our aim is not only to integrate existing records, but above all to share them with experts who are able to improve the quality of the existing records, to fill the gaps in the distribution data, and to link these records to other information on the Czech flora, such as plant traits. This database, launched in 2014, is also serving as a basic platform for mapping the distributions of plants in this country (Kaplan et al. 2015, 2016a, b, 2017a, b, 2018a, b). Here we describe the main sources of the plant occurrence records integrated into the Pladias database and compare different contributing databases in terms of their spatial, temporal and taxonomic coverage. We also describe the methods and workflow for data collation, validation and map preparation.

Methods

Structure and localization of the records

The database model for the species occurrence data within the Pladias database has been kept simple in order to facilitate the integration of data from different sources. A record in its simplest form includes the scientific name of the taxon, locality, date of record, author's name and source of data. Many other fields are optional (Electronic Appendix 1), of which the most important is the status of occurrence (spontaneous, spontaneous secondary, planted or not set). The recording of the locality is point-based using geographic coordinates in the WGS84 coordinate system and has an accuracy given in metres from the indicated location. Less accurate localities based on an assignment to a field of the mapping grid are allowed for old records lacking coordinates. The recording grid consists of cells of 5 longitudinal minutes × 3 latitudinal minutes (approx. 32 km^2 at 50° N latitude), which correspond to quadrants of the basic fields (10 × 6 minutes) of the grid template used for Central European floristic mapping (Niklfeld 1971, 1999, Schönfelder 1999). These basic fields, sometimes referred to as the Central European Basic Area (CEBA; Petřík et al. 2010), were originally defined by the sheets of the German topographic map (Messtischblätter) at a scale of 1: 25,000 (Niklfeld 1971). Different software and projects implemented this grid following the given latitudinal and longitudinal template, but using different coordinate systems. It resulted in spatial differences of up to hundreds of meters between boundary lines in various implementations. To create a grid for the Pladias database, we used a template of the grid implemented in the Janitor software, developed by the Czech Environmental Information Agency (CENIA), which most closely matches the grid used for the localization of old records, mainly in FLDOK. Unfortunately, the Janitor grid does not include quadrants in the neighbouring countries, and some fields were added later to the initial grid, resulting in significant spatial errors. Therefore we used the same definition of basic quadrants and their divisions based on the

Pulkovo 1942 coordinate system (EPSG: 4284) as used in Janitor and created a larger grid covering most of central Europe to accommodate all the records from adjacent areas. The grid was subsequently transformed into WGS84 (EPSG: 4326). The Czech Republic is covered by 2,551 quadrants (quarters of the basic fields), of which 2,181 are entirely within the borders of the country. Maps in the Pladias web application (www.pladias.cz) and other printed outputs are displayed using the spherical Mercator projection (EPSG: 3857), in which meridians and parallels appear as straight lines, and the grid cells are thus displayed as squares.

Taxonomic concept

The species list included in the Pladias database is derived from the hierarchical checklist of the Czech flora published by Danihelka et al. (2012) and the family classification and circumscriptions described in version 11 of the Angiosperm Phylogeny Website (Stevens 2001 onwards), which was an improved version of the APG III classification (Angiosperm Phylogeny Group 2009). For ferns, the assignment of genera to families follows Smith et al. (2006). The original checklist (Danihelka et al. 2012) was further developed by correcting a few technical mistakes, adding recently described species and new records for the Czech flora, and, if necessary, making nomenclatural changes. Before importing any of the large datasets of species occurrences, names of taxa used in this dataset were matched with those in the database, and the matches were subsequently checked and corrected, using the expertise acquired when working with plant records from different periods in the past. The original name under which a record was published or entered in a source database is preserved in a separate database field. When herbarium specimens are recorded, the original identification of the specimen is usually preserved in that field. This should reduce the number of mistakes when records are taxonomically reinterpreted and facilitate record checking in the original sources.

Data handling and validation

Unique features of the Pladias database at the national level are automated data-checking procedures applied during the import, data validation process and the stress on data correctness and accuracy. Records enter the database in two ways. Large data sets from existing databases managed by various institutions (see Table 1 and Appendix 1) are imported in bulk by the managers of the Pladias database. The same procedure will be repeated in the future to import records recently added to these source databases. Small data sets provided by taxonomic experts and collaborators are imported by the contributors themselves as MS Excel spreadsheets with a predefined structure and restricted range of values for several fields. We preferred this simple approach instead of a web form because it does not require an internet connection for filling in the files and because most of the collaborators are familiar with MS Excel. Most of these imports are restricted to a limited number of distribution maps of taxa that were produced in a particular period. Both types of data are carefully checked for formal and content inconsistencies before they are imported into the database. For instance, all location data such as geographic coordinates, the cadastral unit, administrative district, grid cell and phytogeographical district are automatically cross-validated against each other if there are at least two of them for a particular record.

Table 1. – Overview of databases and projects contributing species records to the Pladias database. The total number of records and the total number of taxa in each source and the number of taxa contributed by each source are given (as of 18 December 2018). Czech acronyms of particular databases and projects (if commonly used) are given in brackets.

Database or project	Acronym	No. of taxa	No. of unique taxa	No. of records
Database of the Distribution of Vascular Plants in the Czech Republic	FLDOK	4,192	383	1,809,115
Species Occurrence Database of NCA CR	NDOP	3,721	185	7,500,871
Czech National Phytosociological Database	CNPD (ČNFD)	2,858	17	2,052,886
Taxonomic Experts' Records	Experts	2,701	99	570,414
Flora of Eastern Bohemia	KVC	2,532	24	165,936
Records from the Summer Schools of Field Botany of the Czech Botanical Society	FK CBS	2,106	2	123,541
Floristic Database of the South Bohemian Branch of the Czech Botanical Society	JCP CBS	1,968	4	129,377
Floristic Database of the Moravian-Silesian Branch of the Czech Botanical Society	MSP CBS	1,628	14	21,837
Floristic Database for the Vysočina Region	DKV	1,518	1	23,617
Recording Cards of Bohumil Slavík	Slavík's Cards	1,431	1	136,450
Floristic Archive of Bohumil Slavík	Slavík's Archive	1,344	1	17,560
Database of the Czech Forest Classification System	DCFCS (DLT)	1,336	4	969,759
Flora of the Ještěd Ridge	FJR	1,056	5	58,391

The objective of the Pladias database is not only the integration of data but also continuous improvement of data quality based on data sharing and feedback from scientists involved in the project, volunteers, data administrators and data owners. For this purpose, we developed a graphical user interface, which allows for validation of occurrence data by taxonomic experts and also for checking and commenting on individual records by registered users, usually field botanists experienced in floras of different areas within the country. Original records of a particular species are always displayed for a particular occupied quadrant after clicking on the quadrant in the grid map (see Electronic Appendix 1 for technical description and screenshots). Three main attributes of records can be set by an expert in an overview window, which facilitates the process of validation: (1) record's reliability (correct, incorrect, uncertain or not evaluated yet), (2) herbarium voucher revised (false or true; only for herbarium specimens seen by a taxonomic expert in charge of the particular taxon) and (3) status of occurrence (spontaneous, spontaneous secondary, planted or not set). Most other attributes can be changed by the taxonomic expert in charge of the species, using a more detailed editing form displayed for each individual record (see Electronic Appendix 1 for a list of fields available). Any registered user can add a comment to any record, and this comment is subsequently considered by an expert who can finally tag the comment as processed. All changes in a record are tracked and accessible to users as a list.

Records are validated and edited either during the preparation of distribution maps by taxonomic experts or by the team of database administrators during a targeted search for errors. All edits and comments made in the Pladias database are regularly reported back to the owners or administrators of the active source databases, which results in an improvement in the quality of records in all the databases.

Distribution maps

One of the purposes of the occurrence module within the Pladias database is to create a platform for the preparation of distribution maps based on critically evaluated records. In addition to the validation and visualization of species distribution records, the graphical user interface provides complex functionality for the administration of the individual steps in the mapping procedure (see Electronic Appendix 1). A map draft is prepared by its author assuming that each grid cell mapped as occupied contains at least one validated record, though usually it contains several records. Draft distribution maps are then reviewed by the registered users and their comments are considered by the authors before producing the final maps. Maps are prepared separately using the ArcGIS 10.5 software and Python 2.7. scripts, usually in large batches. Once a map is to be generated, the database creates a backup of all the records to be used in a particular map. The map produced is then combined in one PDF with the list of all records used in this map. Final maps are published in the journal Preslia (www.preslia.cz) as well as on the public web Pladias: Database of the Czech Flora and Vegetation (www.pladias.cz).

Data analyses

The Pladias database is an ongoing project, and new records are continuously imported and evaluated by taxonomic experts. Here we analyse the records integrated in the Pladias database by November 2017. All the results are thus related to this date unless another date is given. All statistical analyses were performed in R version 3.4.1 (R Core Team 2017) using the RPostgreSQL package. The Postgre database was queried in pgAdmin version 1.18.1. All spatial analyses and visualizations were done in ArcGIS 10.5.

Results

Spatiotemporal coverage

In November 2017, the Pladias database contained 13,134,183 records of 4,883 vascular plant taxa (13,579,754 records of 4,921 taxa in December 2018), which came from five large national databases and seven regional projects, supplemented by records collected by taxonomic experts for the Atlas of Distribution of Vascular Plants in the Czech Republic (Taxonomic Experts' Records). The number of records per quadrant (only those quadrants that were entirely within this country's borders were considered, see Fig. 1A) is extremely variable, pointing to large spatial variation in sampling effort. While 10 quadrants have less than 300 records, there are 26 quadrants with more than 25,000 records (minimum = 79, maximum = 54,697, mean = 5,272, median = 3,828). The number of taxa per quadrant (Fig. 1B) varies from 49 (less than 100 taxa were found in only two quadrants) to 1,347 (mean = 567, median = 547). There is a high variance in both variables, though the number of records varies much more (coefficient of variation for records = 1.06 and for taxa = 0.35) and they are only moderately correlated (Spearman's rho = 0.68).

Individual source databases strongly differ in the total number of records and the number of taxa recorded (Fig. 2, Table 1). For instance, NDOP by far surpasses other sources in the number of records, whereas FLDOK contains records of almost 400 taxa not included in any other data source. The targeted effort of experts focused on particular

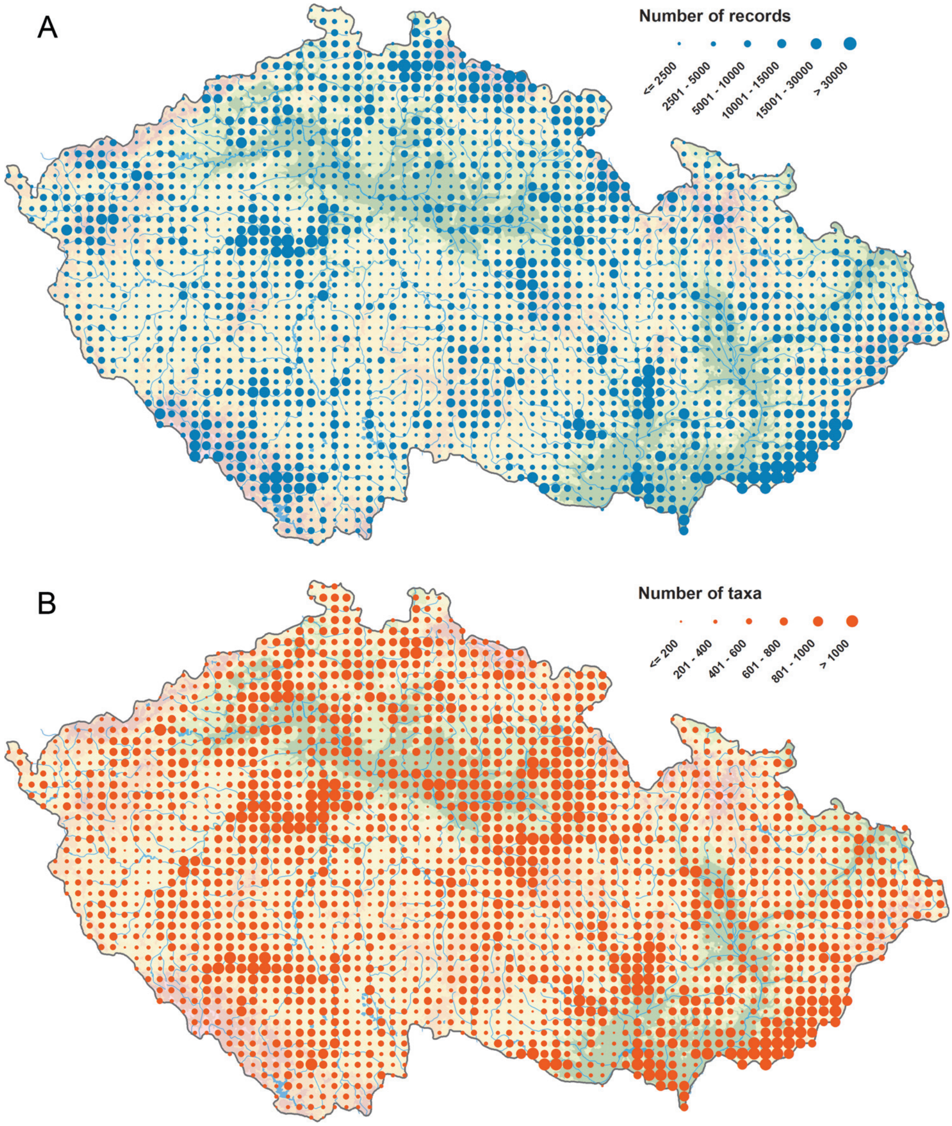

Fig. 1. – Total number of occurrence records (A) and taxa (B) per quadrant in the Pladias database.

species (Taxonomic Experts' Records) added up to more than 15% of the number of taxa occurring in particular quadrants (Fig. 3).

There are also considerable differences among the lists of the most common species in individual databases (Table 2). While the FLDOK's list of the 15 most frequent taxa contains only two species that are not listed among the most frequent taxa of other national databases, NDOP has six and DCFCS even nine such species. High spatial variation in sampling effort is pronounced also at the level of individual source databases. Even the

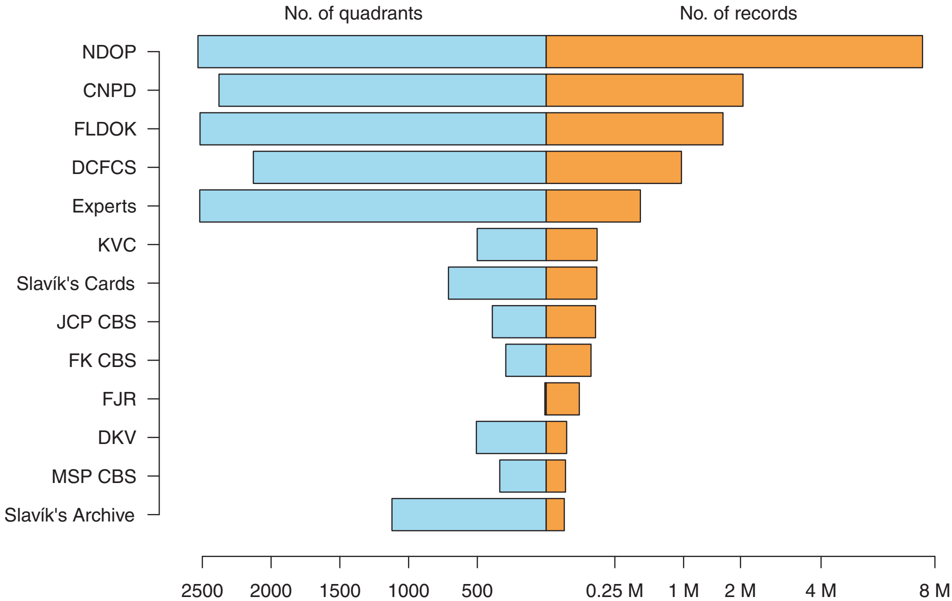

Fig. 2. – Contribution of different sources of data (y-axis) to the total number of occupied quadrants (left) and records (right, log scale) in the Pladias database (x-axis). See Table 1 or Appendix for full names of individual databases.

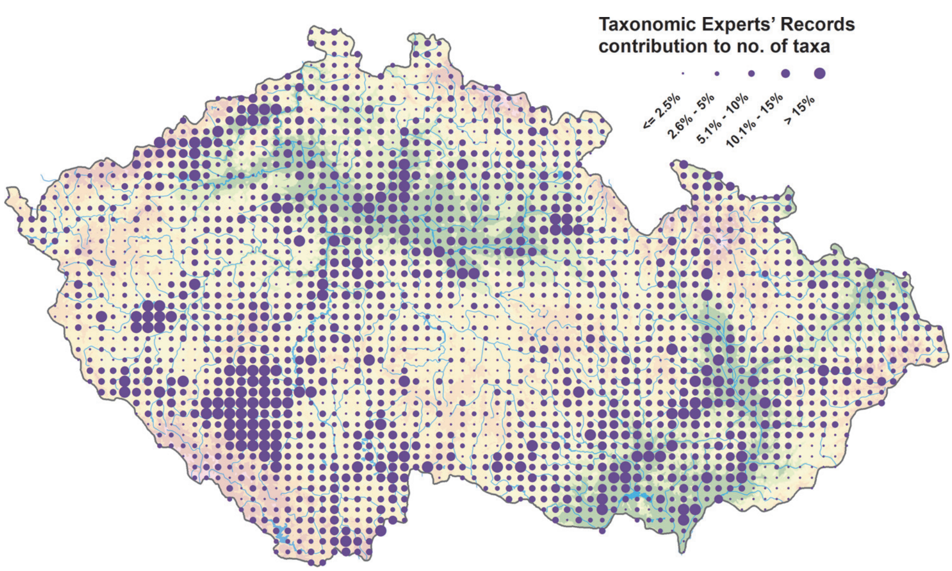

Fig. 3. – Unique contributions of new records collected by taxonomic experts to the total number of taxa in quadrants. Most of these records are based on revised herbarium specimens. Note that these records focused on the 570 taxa whose maps were finished in November 2017.

Table 2. – The first 15 most frequent taxa in the five major data sources based on the number of records. The taxa listed only once are in bold.

FLDOK	NDOP	CNPD	Slavík's Cards	DCFCS
Urtica dioica	*Picea abies*	*Urtica dioica*	*Urtica dioica*	*Picea abies*
Lemna minor	*Urtica dioica*	*Taraxacum* sect. *Taraxacum*	*Achillea millefolium*	*Vaccinium myrtillus*
Juncus effusus	*Fagus sylvatica*	*Dactylis glomerata*	*Veronica chamaedrys*	*Avenella flexuosa*
Picea abies	***Betula pendula***	*Deschampsia cespitosa*	***Heracleum sphondylium***	***Oxalis acetosella***
Ranunculus repens	*Vaccinium myrtillus*	*Ranunculus repens*	*Plantago lanceolata*	*Fagus sylvatica*
Arrhenatherum elatius	***Alnus glutinosa***	*Festuca rubra*	*Dactylis glomerata*	***Sorbus aucuparia***
Deschampsia cespitosa	*Pinus sylvestris*	*Plantago lanceolata*	*Hypericum perforatum*	***Hieracium murorum***
Dactylis glomerata	***Acer pseudoplatanus***	*Picea abies*	***Lotus corniculatus***	*Pinus sylvestris*
Hypericum perforatum	***Fraxinus excelsior***	***Rumex acetosa***	*Ranunculus repens*	***Rubus idaeus***
Euphorbia cyparissias	***Prunus spinosa***	*Arrhenatherum elatius*	***Plantago major***	***Mycelis muralis***
Veronica chamaedrys	*Arrhenatherum elatius*	***Ranunculus acris***	***Knautia arvensis***	*Senecio ovatus*
Vaccinium myrtillus	*Avenella flexuosa*	*Achillea millefolium*	***Aegopodium podagraria***	***Luzula luzuloides***
Festuca rubra	*Dactylis glomerata*	***Elymus repens***	*Euphorbia cyparissias*	***Athyrium filix-femina***
Taraxacum sect. *Taraxacum*	*Deschampsia cespitosa*	*Veronica chamaedrys*	***Sambucus nigra***	***Luzula pilosa***
Agrostis capillaris	***Carpinus betulus***	*Avenella flexuosa*	*Senecio ovatus*	***Abies alba***

largest database (NDOP) shows strong spatial variation in record density, and what is more important, the pattern is different among databases in terms of both the number of records and the taxa recorded (Fig. 4, Electronic Appendix 2).

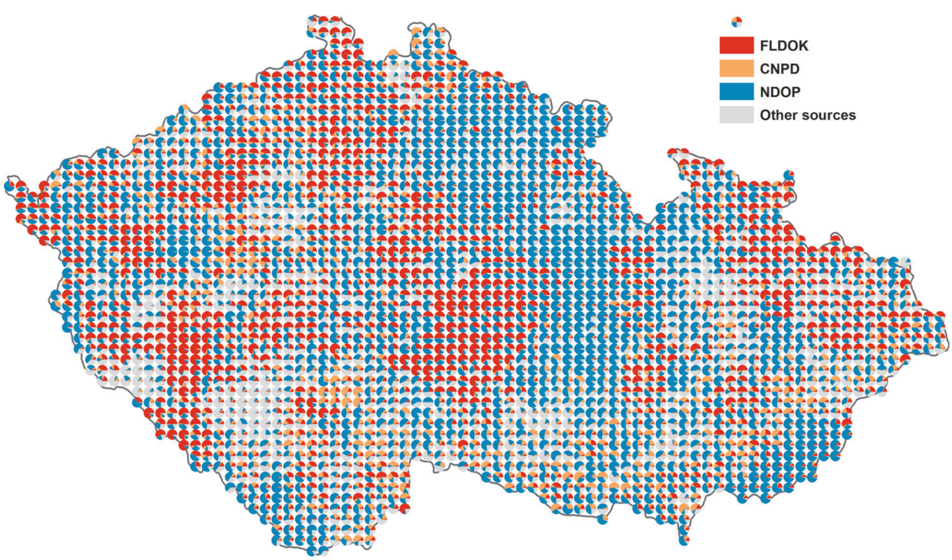

Fig. 4. – Unique contributions of FLDOK, CNPD, NDOP, and other databases and projects together to the total number of taxa in each quadrant. The numbers of taxa present in only one data source and absent from the rest of the Pladias data were counted for each particular quadrant. Proportions of the unique taxa from the three largest source databases (FLDOK, CNPD and NDOP) and the sum of unique taxa from the other sources are displayed.

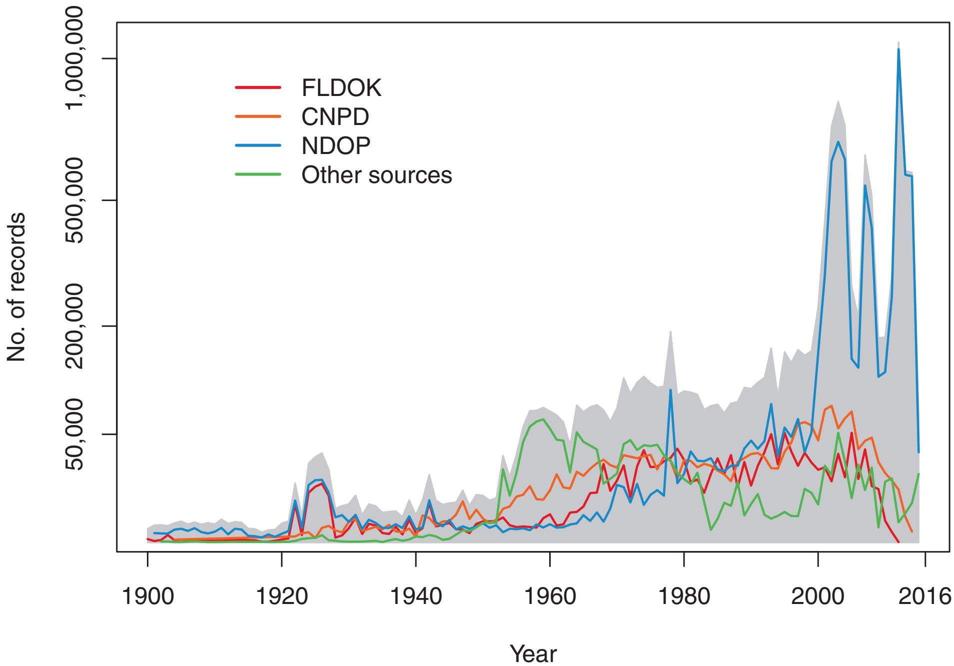

Fig. 5. – Number of plant occurrence records according to the year of collection in the Pladias database. Records before 1900 (0.1% of the total) were omitted. Three largest data sources (FLDOK, CNPD and NDOP), other sources and the sum of all sources (grey silhouette) are displayed. Note the square-root transformation of the y-axis.

There is a marked contrast in the small number of records from the pre-1950 period and the vast number of new records in the last decade (Fig. 5). Only about 900 records are from the period before the year 1800, and only 0.1% of the total amount is from the period before 1900 (Fig. 5). In contrast, the records collected after 2000 account for about 60% of the total and most of them were collected during the habitat mapping project organized by the NCA CR. Apparent accumulations of records in some decades can be attributed to particular activities. For instance, most of the records from the 1920s were collected by S. Staněk, an amateur botanist working in the Bílé Karpaty Mts in the southeast of the country (Staněk et al. 1996), while the peak around the 1960s is related to forest vegetation sampling organized by the Forest Management Institute (database DCFCS).

A comparison of the taxonomic coverage of the Pladias database and the national plant checklist

Present records include 82% of the taxa in the national vascular plant checklist (Danihelka et al. 2012). Taxa for which there are no records (n = 878) may be divided into six groups. A large group comprises hybrids (n = 243) or other taxa difficult to identify, such as *Taraxacum* or *Hieracium* species (n = 123), which are often not identified to the species level during ordinary field surveys. Another group includes taxa distinguished in the national checklist at lower taxonomic ranks (subspecies, variety and form; n = 95) or higher ranks (aggregates and groups; n = 28). Of these, either only a taxon at a higher rank (usually a species) is recorded instead of its type subspecies or variety, which is the only subspecies present in the Czech flora (e.g. *Alnus incana* was recorded instead of *A. incana* subsp. *incana*), or the distribution of a taxon at the respective rank is unknown. Names of many crops and cultivated ornamental plants are present in this database for practical and technical reasons but they are not recorded (n = 169; see Kaplan et al. 2015). The rest of the names lacking any record (n = 222) refers to recently introduced species or rare casuals (e.g. *Ambrosia psilostachya*) or to taxa whose past occurrence in the Czech Republic is uncertain (e.g. *Carex brevicollis*).

Data quality

Producing of the distribution maps for the first 570 taxa included in the mapping project, taxonomic experts had validated almost 170,000 records by the end of 2017. If we also count the records that were imported as a dataset Taxonomic Experts' Records, which we consider as correct if the importer is in charge of a particular species' map, the number of validated records is approximately 0.5 million (Table 3). Despite the huge effort of all the experts, the validated records form only a small percentage of the total number of records; 1.3% or 3.9%, without and with the Taxonomic Experts' Records, respectively. Of the ~170,000 validated records, 83.6% were classified as correct (but sometimes only after correcting apparent mistakes in geographic location), 5.2% were classified as incorrect and 11.2% as uncertain. Most of the original sources are similar in the percentage of correct, incorrect and uncertain records. The lowest share of records evaluated as correct was found in the project Recording Cards of Bohumil Slavík, which is mainly due to the poor metadata; usually only the taxon name and quadrant are recorded. Many records from this dataset were thus classified as uncertain (28.8%). The NDOP database contained the highest percentage of records classified as incorrect (7%).

Table 3. – Total number of validated records in each source database or project, their percentage share of the total number of validated records and of records marked as correct, incorrect or uncertain among the validated records from a particular source. For full names of individual databases or projects see Table 1. Please note that the percentage of correct records is slightly overestimated because for some of these records geographic location was corrected before marking them as correct.

Database or project	Number of validated records	% of the total number of records	% of the number of validated records		
			Correct	Incorrect	Uncertain
FLDOK	48,434	3.0	88.0	4.2	7.8
NDOP	72,005	1.0	78.4	7.0	14.6
CNPD (ČNFD)	32,564	1.6	88.4	3.6	8.0
Experts	341,472	72.8	99.7	0.1	0.2
FK CBS	2,833	2.7	92.3	0.4	7.3
JCP CBS	2,916	2.3	92.0	2.0	6.0
MSP CBS	674	3.4	91.8	0.3	7.9
DKV	1,256	5.6	90.0	2.8	7.2
Slavík's Cards	2,887	2.1	68.0	3.2	28.8
Slavík's Archive	391	2.2	88.0	2.8	9.2
DCFCS (DLT)	4,563	0.5	79.8	6.7	13.5
FJR	232	0.4	85.0	0.0	15.0
Total	510,229	3.9	94.3	1.8	3.9
Total without Experts	168,757	1.3	83.6	5.2	11.2

Discussion

Advantages and challenges of integrating plant distribution data at a national scale

The new Pladias database contains the largest set of data on the occurrence of vascular plants in the Czech Republic, which is subject to continuous quality control. With approximately 13.6 million species occurrence records, it is one of the largest databases of this kind globally. For example, the GBIF database (www.gbif.org, accessed in December 2018) contains more records only for five countries (France, UK, Australia, Germany and the Netherlands), most of which are much larger than the Czech Republic. If databases for neighbouring countries are considered, only the German database with its more than 30 million records (Netzwerk Phytodiversität Deutschland & Bundesamt für Naturschutz 2013) contains more than twice the number of records in the Pladias database. The Polish database ATPOL (Zając & Zając 2002) currently contains about 6.5 million records (Marcin Nobis, in litt.), and the Austrian database 1.95 million (University of Vienna, Faculty of Life Sciences, Department of Botany and Biodiversity Research, Research Group Plant Biogeography 2018).

Large databases of species occurrence records coming from different sources, ranging from local projects to national recording schemes, are increasingly used for testing various ecological hypotheses (e.g. Kühn et al. 2006, Whittaker et al. 2007, Svenning et al. 2008, Ronk et al. 2017). However, the results of such projects are dependent on the quality, representativeness and comprehensiveness of the data. Based on the comparison of all the data sources listed in Table 1, it is clear that even a small local project may contribute records of taxa not included in any other database. Not only the detectability of the rarest species, but also estimates of the abundance of the most common species, are dependent on the quality of the data. Indeed our analysis shows that the estimates of the most abundant species in this country obtained using the individual source databases differ. Thus the most important advantage of merging data from different sources is that these sources are partly complementary and mutually compensate for missing records or taxa in some areas.

Even in a large database that integrates millions of species occurrence records from all major national sources, however, it is still unclear to what extent the emerging biogeographical and ecological patterns are influenced by various biases in the data. The observed uneven distribution of species records is partly due to the spatial distribution of species richness, but sampling effort also contributes significantly to this variation. It is apparent that some areas, such as the strip of intensively used farmland south-west of Brno, are under-sampled, while other areas, such as the surroundings of Prague and some parts of southern Moravia, are over-sampled (Fig. 1). The reason for oversampling is usually a combination of high attractiveness of their floras and a high number of botanists working in the area. Thus merging data from different sources is more likely to indicate where under- and over-sampled areas occur, and so enable the planning of future sampling activities so that they contribute to gap-filling. Nevertheless even the combination of all existing data still does not cover the occurrence of all the taxa reported in this country. Targeted extractions of herbarium records and literature sources together with local field surveys during the first four years of the PLADIAS project added a substantial number of taxa occurring in particular quadrants (Fig. 3), but so far this only accounts for a small part of the national checklist. It is clear that significant effort is still necessary to use the majority of the already existing records to produce reliable distribution data.

Any project integrating species occurrence data faces various challenges, of which duplications are among the most important. Electronic databases often contain duplicated or multiplicated records independently extracted from the same original source. For example, a record of a rare endangered plant originally recorded in a vegetation plot (relevé) could be entered in all the largest original databases (FLDOK, NDOP and CNPD) and then imported into the Pladias database, thereby triplicating a single find. Although we tried to remove the most striking duplicate subsets before merging data, double or multiple entries are common and even visible as synchronized local maxima along the time axis among data sets. For instance, the striking increase in the number of records in the 1920s (Fig. 4) is caused not only by the high recording activity of a local botanist (S. Staněk), but also by duplication of records, which were included in the FLDOK and the NDOP databases. However, dupli- and multiplicates are also useful for detecting spatial and other errors by comparing the corresponding records when validating the data.

Data integration inevitably leads to the loss of some information. As data from all sources have to be transformed into a uniform database structure, some information can be incompatible and has to be reduced or even omitted. For instance, in our case, the localizations as lines or polygons used in the NDOP database had to be transformed to the centroid of these objects to be compatible with the point-based localization system used in the Pladias database.

Data quality

Generally, all databases contain errors, which are multiplied by integrating databases developed for different purposes and different in terms of quality. Many errors can be removed during record validation by expert botanists with local field experience, but this process is time-consuming and resource-demanding. Despite joint efforts of many experts involved in the PLADIAS project and external collaborators, so far only a small percentage of the records imported into the Pladias database have been validated. However, the absolute number of these records is more than 0.5 million, which allowed the production of 674 distribution maps by December 2018 (Kaplan et al. 2018b) based on revised records and the evaluation of the quality of the data and prevailing type of errors in each original database or project. We have shown that there are no big differences between source databases and projects in terms of percentage of records classified by experts as correct, incorrect or uncertain, but they differ in the prevailing types of errors. Below we describe the various types of errors that we dealt with:

(1) Correct identification of a plant depends on the knowledge and field experience of individual botanists, which varies considerably among people and taxonomic groups. It is particularly difficult to identify errors in the identification of common species. In rare species or species with a specific ecology, such errors can be eliminated but it requires substantial taxonomic and floristic expertise. *Cerastium lucorum* can serve as an example of the validation process in the Pladias database and critical assessment of records by experts. This species is frequently confused with the common *C. holosteoides* but differs in its ecology (Smejkal 1990). Based on this knowledge and a distribution map of *C. lucorum* prepared solely based on revised herbarium specimens, the occurrences outside the geographic range of this species and occurrences in habitats in which this species

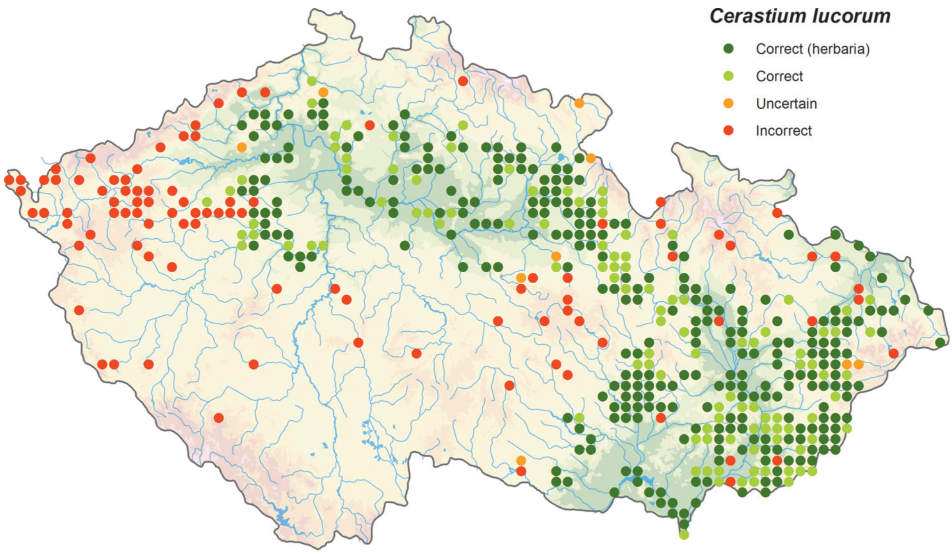

Fig. 6. – An example of how critical evaluation can reveal a high number of erroneous records. Of the quadrants for which the occurrence of *Cerastium lucorum* is recorded in the Pladias database 28% were assessed as uncertain or even incorrect.

normally does not occur were tagged as incorrect. The number of occupied quadrants after validation was reduced by 28% (Fig. 6).

(2) Errors in extracting information from published sources. Various technical errors can occur during computer data processing or record typing, for example, unwanted shifting of rows in tables (sometimes already present in printed sources) and lists or shuffling site codes. Specific mistakes are produced when technicians without a botanical background computerise written records for the source databases. These are generally difficult to detect unless rare species or species with specific ecology are concerned. Specific types of errors occur in databases in which abbreviations of a species names are selected from a standard list during the entry of a record rather than typing whole plant names; examples include confusions between *Carex pilulifera* and *Carex pilosa* if the code "Car pil" is used, or among *Calla palustris*, *Callitriche palustris* and *Caltha palustris* ("Cal pal"). Even confusion between bryophyte and vascular plant species occasionally occur (e.g. "Cal str" for *Calliergon stramineum* and *Calamagrostis stricta*). These errors are especially typical of the species records from vegetation plots that were first entered into the Turboveg database (Hennekens & Schaminée 2001). However, similar mistakes are also frequent in the FLDOK or the NDOP databases, in which names are selected from a list.

(3) Location errors. These concern mainly misinterpretations of the same names used for several different places in a country, old records with German place names, or obsolete place names that are no longer in use. Again, most of these errors are detectable for rare species. Many location errors, mainly typing mistakes, are detected by the automated check during the import of data if at least two different geographical attributes are given for a record (e.g. quadrant code and coordinates, see Methods). Otherwise, the chance of identifying location errors is much lower, except for obvious outliers on distribution

maps. However, in the case of vegetation-plot data, if a location error is revealed for one species by detecting it as an outlier on a species distribution map, this error is subsequently corrected for all the records from the same plot.

Herbarium databases as a possible source of historical records

The paucity of historical records can be partly reduced by including herbarium databases, which is, however, a difficult task. In general, three different electronic systems are widely used for herbarium databases in this country, and only the herbarium databases of Masaryk University and the Moravian Museum in Brno each contain more than 100,000 records of vascular plant specimens, of which about 70% are from the Czech Republic. A pilot analysis has shown that the quality of the records in these databases differs greatly and many records lack some of the important attributes, for instance reliable geographic coordinates or the original text on the label. A central import of some herbarium databases, which may be regarded as the most efficient solution, would not only be a very laborious operation, but it would inevitably result in many erroneous and duplicate records. However, the absolute number of old records gradually increases with the import of records based on herbarium specimens revised by taxonomic specialists.

Feedback to source databases

Probably the most challenging task in the management of an integrated database of plant distribution records is the development of an efficient way of tracking changes made in different source databases and the Pladias database and the mutual exchange of this information. Establishing a live link between the databases would enable the efficient management of changes, reducing duplication and propagation of errors. The best solution for database management would be a single central database of species' distribution records developed and used jointly by all the involved institutions. However, such a development is very unlikely because of the substantial differences between the databases, their future plans, the purpose of these databases and their legal status. The approach proposed by the Pladias database seems to be the best feasible solution for Czech botany at the moment.

Future outlook

The current explorative analysis of the plant distribution records in the Pladias database showed the great potential of this project for significantly improving our knowledge of plant distribution. This has been demonstrated by a series of distribution maps produced by a joint effort of taxonomic experts and regional botanists, which make the production of a complete atlas of the distribution of vascular plants in the Czech Republic both viable and feasible.

The thorough inspection of the database also suggests several ways of improving the quality of the data and spatial coverage in the near future. More errors can be detected and corrected by using new automatic and semi-automatic procedures for controlling the quality. For instance, geographic coordinates can be checked against the altitude of the site derived from precise digital elevation models. Similarly, building a unified list of source literature improves searching for duplicates and facilitates further additions to the database.

However, building a plant distribution database is not the ultimate goal. The distribution data can be used for a variety of analyses in the context of both academic and applied research. In particular, linking these data to data on species traits and environmental conditions, and the integration of all this information into international projects is very promising.

See www.preslia.cz for Electronic Appendices 1–2

Acknowledgements

Development of the plant occurrence module of the Pladias database would not be possible without the efforts of innumerable botanists who recorded data in the field and contributed to particular databases (see the list of contributors below). Data integration and related software development were funded by the Centre of Excellence PLADIAS, project no. 14-36079G from the Czech Science Foundation. JB, JD, JŠ, JW, PP and ZK were also supported by the long-term research development project no. RVO 67985939 from the Czech Academy of Sciences. Tony Dixon kindly improved our English.

List of registered and active individual contributors of plant occurrence data to the Pladias database

Martina Bartošová, Jiří Bělohoubek, Jan Blahovec, Jiří Brabec, Leo Bureš, Petr Bureš, Roman Businský, Luděk Čech, Alžběta Čejková, Martin Čertner, Jindřich Chrtek, Martin Dančák, Jiří Danihelka, Jan Doležal, Pavel Dřevojan, Michal Ducháček, Martin Duchoslav, Václav Dvořák, Libor Ekrt, Ester Ekrtová, Karel Fajmon, Irena Formanová, Hana Galušková, Michal Gerža, Vít Grulich, Miloš Hájek, Rudolf Hlaváček, David Hlisnikovský, David Horák, Ondřej Hornych, Věra Hromádková, Michal Hroneš, Zdenka Hroudová, Jana Janáková, Boleslav Jelínek, Jaroslav Jirásek, Petra Juřáková, Jiří Juřička, Nikol Kantorová, Zdeněk Kaplan, Jan Kirschner, Lucie Kobrlová, Jiří Kocián, Petr Kocián, Josef Komárek, Jan Košnar, Milan Kotilínek, Daniel Koutecký, Petr Koutecký, František Krahulec, Karel Kubát, Josef Kučera, Pavel Kúr, Martin Lepší, Petr Lepší, Pavel Lustyk, Jiří Malíček, Vladimír Melichar, Jana Möllerová, Patrik Mráz, Zuzana Mruzíková, Pavel Novák, Klára Nunvářová Kabátová, Radim Paulič, Tomáš Peterka, Petr Petřík, Lenka Pivoňková, Jan Prančl, Romana Prausová, Olga Rotreklová, Zdenka Rozbrojová, Radomír Řepka, Věra Samková, Milan Skalický, Hana Skálová, Ivan Suchara, Tomáš Svačina, Lenka Šafárová, Ota Šída, Lenka Šoltysová, Martin Šrámek, Martin Štastný, Milan Štech, Radek Štencl, Jitka Štěpánková, Petra Štěpánková, Kateřina Šumberová, Přemysl Tájek, Vojtěch Taraška, Tomáš Tichý, Bohumil Trávníček, Radim J. Vašut, Michal Vávra, Jiří Velebil, Jaroslav Zámečník, Jiří Zázvorka.

Souhrn

Výzkum květeny má v České republice dlouhou tradici, a přinesl proto obrovské množství lokalizovaných záznamů o rozšíření cévnatých rostlin. Během posledních tří desetiletí vzniklo nezávisle na sobě mnoho elektronických databází, do nichž se tyto údaje ukládají nebo ukládaly a které jsou spravovány různými institucemi, jejichž pracovníci jsou zastoupeni v autorském kolektivu tohoto přehledového článku. Jsou to mj. databáze FLDOK (Floristická dokumentace) spolu se škrtacími seznamy a archivem B. Slavíka (vše Botanický ústav AV ČR), ČNFD (Česká národní fytocenologická databáze, Ústav botaniky a zoologie Přírodovědecké fakulty Masarykovy univerzity), NDOP (Nálezová databáze ochrany přírody, Agentura ochrany přírody a krajiny ČR), Databáze lesnické typologie (Ústav pro hospodářskou úpravu lesů Brandýs nad Labem), které spolu s další databázemi s převážně regionální působností pokrývají všech 2181 kvadrantů síťového mapování (každý o rozloze asi 32 km^2), jež se celou svou plochou nacházejí na území České republiky. Dosavadní roztříštěnost spolu s osobitými charakteristikami každé takové databáze silně omezovaly využití a analýzy údajů o rozšíření rostlin. Východisko ukázal projekt výzkumného centra PLADIAS (**Pla**nt **Di**versity **A**nalysis and **S**ynthesis Centre), financovaný Grantovou agenturou České republiky v letech 2014–2018 v rámci programu na podporu excelence v základním výzkumu. Centrum sdružilo badatele zabývající se diverzitou flóry a vegetace z brněnské Masarykovy univerzity, Botanického ústavu AV ČR a Jihočeské univerzity v Českých Budějovicích. K hlavním cílům projektu patřil výzkum druhové, fylogenetické a funkční diverzity současných flór, jakož i příčin jejího vzniku, dále výzkum procesů utváření rostlinných společenstev a rostlinných invazí. Projekt PLADIAS

také dal název centrální databázi kriticky revidovaných údajů o české flóře a vegetaci, která je veřejně přístupná na portálu pladias.cz.

V tomto článku popisujeme strukturu databáze Pladias, zpracování a validaci dat, přípravu map rozšíření cévnatých druhů založených na kriticky vyhodnocených záznamech (do konce roku 2018 jich vyšlo v časopise Preslia 674) a stručně také zdrojové databáze. Uvádíme výsledky základní analýzy časoprostorového a taxonomického pokrytí sdružených dat o výskytu rostlin. Dosud bylo shromážděno více než 13 milionů záznamů pro téměř 5000 taxonů. Tyto záznamy pocházejí z pěti velkých národních databází a sedmi regionálních projektů. Kromě toho byla část záznamů shromážděna přímo v rámci projektu PLADIAS. I přes nerovnoměrné pokrytí, které je zčásti dáno přírodními podmínkami a lokálně i nedostatečným poznáním květeny, se databáze Pladias stala největším datovým souborem o výskytu cévnatých rostlin v České republice, který podléhá trvalé kontrole kvality. Analýzy integrovaných záznamů ukázaly rozdíly v prostorovém a taxonomickém pokrytí zdrojových datových souborů, a tím prokázaly význam integrace dat, neboť výsledná databáze výrazně zlepšila naše znalosti o rozšíření rostlin. Nejstarší relevantní údaje o květeně českých zemí pocházejí z konce 18. století, jejich množství je však v porovnání s dnešním intenzitou výzkumu téměř zanedbatelné. Floristických údajů začalo výrazně přibývat až v polovině 20. století a jejich množství rychle vzrostlo po roce 2000 s příchodem celonárodních mapovacích projektů se zapojením široké veřejnosti (např. ukončený projekt florabase.cz s 800 registrovanými uživateli) a digitalizací literárních i nepublikovaných údajů. Dosavadní výsledky ukazují, že sběr údajů od botanicky zainteresované veřejnosti je efektivní teprve ve spolupráci s odborníky zaměřenými na validaci existujících záznamů.

References

Angiosperm Phylogeny Group (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. – Botanical Journal of the Linnean Society 161: 105–121.

Beck J., Böller M., Erhardt A. & Schwamghart W. (2014) Spatial bias in the GBIF database and its effect on modelling species' geographic distributions. – Ecological Informatics 19: 10–15.

Brabec E. (1999) Metodika počítačového zpracování [Methodology of computer processing]. – In: Kolbek J., Mladý F. & Petříček V. (eds), Květena CHKO a BR Křivoklátsko. 1. Mapy rozšíření cévnatých rostlin [Flora of the Křivoklátsko Landscape Protected Area and Biosphere Reserve. 1. Distribution maps of vascular plants], p. 7–10, Botanický ústav AV ČR, Průhonice & Agentura ochrany přírody a krajiny ČR, Praha.

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 (in press).

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., Işık 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. & Michalcová D. (2012) Czech National Phytosociological Database. – In: Dengler J., Oldeland J., Jansen F., Chytrý M., Ewald J., Finckh M., Glöckler F., Lopez-Gonzalez G., Peet R. K. & Schaminée J. H. J. (eds), Vegetation databases for the 21st century, Biodiversity & Ecology 4: 345.

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

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

Danihelka J., Chytrý M., Kučera J. & Palice Z. (2017) History of botanical research in the Czech Republic. – In: Chytrý M., Danihelka J., Kaplan Z. & Pyšek P. (eds), Flora and vegetation of the Czech Republic, p. 25–87, Springer, Cham.

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.

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

Franklin J., Serra-Diaz J. M., Syphard A. D. & Regan H. M. (2016) Big data for forecasting the impacts of global change on plant communities. – Global Ecology and Biogeography 26: 6–17.

GBIF (2015) GBIF Annual Report 2014. – Copenhagen. URL: http://www.gbif.org/resource/annual_report_2014 (accessed 19 January 2017).

Graham C. H., Elith J., Hijmans R. J., Guisan A., Townsend Peterson A., Loiselle B. A. & The Nceas Predicting Species Distributions Working Group (2008) The influence of spatial errors in species occurrence data used in distribution models. – Journal of Applied Ecology 45: 239–247.

Grulich V. (1997) Atlas rozšíření cévnatých rostlin Národního parku Podyjí/Thayatal [Distribution atlas of vascular plants of the Podyjí/Thayatal National Park]. – Masarykova univerzita, Brno.

Haeupler H. E. (2005) Ein Beitrag zur Geschichte der floristischen Kartierungen der Farn-und Blütenpflanzen in Mitteleuropa. – Hoppea 66: 133–152.

Härtel H., Lončáková J. & Hošek M. (eds) (2009) Mapování biotopů v České republice: východiska, výsledky, perspektivy [Habitat mapping in the Czech Republic: foundations, results, perspectives]. – AOPK ČR, Praha.

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.

Jalas J. & Suominen J. (eds) (1972) Atlas Florae Europaeae. Distribution of vascular plants in Europe. 1. *Pteridophyta* (*Psilotaceae* to *Azollaceae*). – The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo, Helsinki.

Jongepier J. W. & Pechanec V. (eds) (2006) Atlas rozšíření cévnatých rostlin CHKO Bílé Karpaty [Distribution atlas of vascular plants of the White Carpathians Protected Landscape Area]. – ZO ČSOP Bílé Karpaty, Veselí nad Moravou.

Kaplan Z. (2012) Flora and phytogeography of the Czech Republic. – Preslia 84: 505–573.

Kaplan Z., Danihelka J., Chrtek J. Jr., Prančl J., Ducháček M., Ekrt L., Kirschner J., Brabec J., Zázvorka J., Trávníček B., Dřevojan P, Šumberová K., Kocián P., Wild J. & Petřík P. (2018a) Distributions of vascular plants in the Czech Republic. Part 7. – Preslia 90: 425–531.

Kaplan Z., Danihelka J., Koutecký P., Šumberová K., Ekrt L., Grulich V., Řepka R., Hroudová Z., Štěpánková J., Dvořák V., Dančák M., Dřevojan P. & Wild J. (2017a) Distributions of vascular plants in the Czech Republic. Part 4. – Preslia 89: 115–201.

Kaplan Z., Danihelka J., Lepší M., Lepší P., Ekrt L., Chrtek J. Jr., Kocián J., Prančl J., Kobrlová L., Hroneš M. & Šulc V. (2016a) Distributions of vascular plants in the Czech Republic. Part 3. – Preslia 88: 459–544.

Kaplan Z., Danihelka J., Štěpánková J., Bureš P., Zázvorka J., Hroudová Z., Ducháček M., Grulich V., Řepka R., Dančák M., Prančl J., Šumberová K., Wild J. & Trávníček B. (2015) Distributions of vascular plants in the Czech Republic. Part 1. – Preslia 87: 417–500.

Kaplan Z., Danihelka J., Štěpánková J., Ekrt L., Chrtek J. Jr., Zázvorka J., Grulich V., Řepka R., Prančl J., Ducháček M., Kúr P., Šumberová K. & Brůna J. (2016b) Distributions of vascular plants in the Czech Republic. Part 2. – Preslia 88: 229–322.

Kaplan Z., Danihelka J., Šumberová K., Chrtek J. Jr., Rotreklová O., Ekrt L., Štěpánková J., Taraška V., Trávníček B., Prančl J., Ducháček M., Hroneš M., Kobrlová L., Horák D. & Wild J. (2017b) Distributions of vascular plants in the Czech Republic. Part 5. – Preslia 89: 333–439.

Kaplan Z., Koutecký P., Danihelka J., Šumberová K., Ducháček M., Štěpánková J., Ekrt L., Grulich V., Řepka R., Kubát K., Mráz P., Wild J. & Brůna J. (2018b) Distributions of vascular plants in the Czech Republic. Part 6. – Preslia 90: 235–346.

Kocián P. & Hlisnikovský D. (eds) (2014) Nálezová databáze Moravskoslezské pobočky ČBS a záznamy rostlin během pilotního projektu [Species occurrence database of the Moravian-Silesian Branch of the Czech Botanical Society and plant records during the pilot project]. – Zprávy Moravskoslezské pobočky ČBS 3, Příloha 2: 3–12.

Kolbek J., Mladý F. & Petříček V. (eds) (1999) Květena CHKO a BR Křivoklátsko. 1. Mapy rozšíření cévnatých rostlin [Flora of the Křivoklátsko Landscape Protection Area and National Park. 1. Distribution maps of vascular plants]. – Botanický ústav AV ČR, Průhonice & Agentura ochrany přírody a krajiny ČR, Praha.

Krahulec F. (2012) History of the studies of flora and vegetation in the Czech Republic. – Preslia 84: 397–426.

Kubát K., Hrouda L., Chrtek J. jun., Kaplan Z., Kirschner J. & Štěpánek J. (eds) (2002) Klíč ke květeně České republiky [Key to the flora of the Czech Republic]. – Academia, Praha.

Kühn I., Bierman S. M., Durka W. & Klotz S. (2006) Relating geographical variation in pollination types to environmental and spatial factors using novel statistical methods. – New Phytologist 172: 127–139.

Meyer C., Weigelt P. & Kreft H. (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. – Ecology Letters 19: 992–1006.

Netzwerk Phytodiversität Deutschland & Bundesamt für Naturschutz (eds) (2013) Verbreitungsatlas der Farn- und Blütenpflanzen Deutschlands. – Landwirtschaftsverlag, Münster.

Niklfeld H. (1971) Bericht über die Kartierung der Flora Mitteleuropas. – Taxon 20: 545–571.

Niklfeld H. (1999) Mapping the flora of Austria and the Eastern Alps. – Revue Valdôtaine d'Histoire Naturelle, Suppl., 51: 53–62.

Petřík P. (2006a) Zdroje floristických údaju v České republice: jak je získat, zpracovat a využít [Sources of floristic data in the Czech Republic: how to acquire, process and use them]. – Zprávy České botanické společnosti 41: 309–329.

Petřík P. (2006b) Floristické mapování a jeho využití na různých prostorových škálách [Grid mapping of flora and its employment at various spatial scales]. – PhD thesis, Knihovna Katedry botaniky PřF UK v Praze.

Petřík P., Pergl J. & Wild J. (2010) Recording effort biases the species richness cited in plant distribution atlases. – Perspectives in Plant Ecology, Evolution and Systematics 12: 57–65.

Podhajská Z. & Škapec L. (1983) Technický projekt NATURIS [Technical project NATURIS]. – Ms., Státní ústav památkové péče a ochrany přírody, Praha.

Preston C. D. (2013) Following the BSBI's lead: the influence of the Atlas of the British flora, 1962–2012. – New Journal of Botany 3: 2–14.

R Core Team (2017) R: A Language and Environment for Statistical Computing. – R Foundation for Statistical Computing, Vienna, Austria.

Ronk A., Szava-Kovats R., Zobel M. & Pärtel M. (2017) Observed and dark diversity of alien plant species in Europe: estimating future invasion risk. – Biodiversity and Conservation 26: 899–916.

Růžička I., Charvátová E., Vrzáková M., Dvořáčková K. & Juřička J. (1968–2011) Floristická kartotéka Muzea Vysočiny Jihlava [Floristic card file of the Vysočina Muzeum Jihlava]. – Ms., Muzeum Vysočiny Jihlava, Jihlava.

Schönfelder P. (1999) Mapping the flora of Germany. – Acta Botanica Fennica 162: 43–53.

Slavík B. (1971) Metodika síťového mapování ve vztahu k připravovanému fytogeografickému atlasu ČSR [Methods of grid mapping in relation to the Phytogeographical Atlas of the Czech Socialist Republic, currently under preparation]. – Zprávy Československé botanické společnosti 6: 55–62.

Slavík B. (ed.) (1986) Fytokartografické syntézy ČSR [Phytocartographical syntheses of the Czech Socialistic Republic]. Vol. 1. – Botanický ústav ČSAV, Průhonice.

Slavík B. (ed.) (1990) Fytokartografické syntézy ČR [Phytocartographical syntheses of the Czech Republic]. Vol. 2. – Botanický ústav ČSAV, Průhonice.

Slavík B. (ed.) (1998) Phytocartographical syntheses of the Czech Republic. Vol. 3. – Academia, Praha.

Smejkal M. (1990) *Cerastium* L. – rožec. – In: Hejný S. & Slavík B. (eds), Květena České republiky [Flora of the Czech Republic] 2: 136–151, Academia, Praha.

Smith A. R., Pryer K. M., Schuettpelz E., Korall P., Schneider H. & Wolf P. G. (2006) A classification for extant ferns. – Taxon 55: 705–731.
Staněk S., Jongepierová I. & Jongepier J. W. (1996) Historická květena Bílých Karpat [Historic flora of the White Carpathians]. – Sborník Přírodovědného klubu v Uherském Hradišti, Suppl. 1: 1–194.
Štěpánková J. (ed.) (2012) Phytocartographical syntheses of the Czech Republic. Vol. 4. – Institute of Botany, ASCR & Academia, Praha.
Stevens P. F. (2001 onwards) Angiosperm PhylogenyWebsite. Version 11, 2011 [and more or less continuously updated since]. – URL: http://www.mobot.org/MOBOT/research/APweb (accessed 19 January 2017).
Svenning J.-C., Normand S. & Kageyama M. (2008) Glacial refugia of temperate trees in Europe: insights from species distribution modelling. – Journal of Ecology 96: 1117–1127.
University of Vienna, Faculty of Life Sciences, Department of Botany and Biodiversity Research, Research Group Plant Biogeography (2018) Floristische Kartierung Österreichs – Mapping the flora of Austria, Occurrence dataset. – URL: https://doi.org/10.15468/eck8bq (accessed via GBIF.org on 18 March 2018).
Westhoff V. & van der Maarel E. (1973) The Braun-Blanquet approach. – In: Whittaker R. H. (ed.), Ordination and classification of communities, p. 615–726, W. Junk, the Hague.
Whittaker R. J., Nogués-Bravo D. & Araújo M. B. (2007) Geographical gradients of species richness: a test of the water-energy conjecture of Hawkins et al. (2003) using European data for five taxa. – Global Ecology and Biogeography 16: 76–89.
Yang W., Ma K. & Kreft H. (2013) Geographical sampling bias in a large distributional database and its effects on species richness–environment models. – Journal of Biogeography 40: 1415–1426.
Zając A. & Zając M. (eds) (2002) Atlas rozmieszczenia roślin naczyniowych w Polsce [Distribution atlas of vascular plants in Poland]. – Pracowńa chorologii komputerowej Instytutu Botaniki Uniwersytetu Jagiellońskiego, Kraków.
Zouhar V. (2012) Database of Czech Forest Classification System. – In: Dengler J., Oldeland J., Jansen F., Chytrý M., Ewald J., Finckh M., Glöckler F., Lopez-Gonzalez G., Peet R. K. & Schaminée J. H. J. (eds), Vegetation databases for the 21st century, Biodiversity & Ecology 4: 346.

Received 22 August 2018
Revision received 27 December 2018
Accepted 28 December 2018

Appendix 1. – Brief descriptions of the largest databases (except for Taxonomic Experts' Records) that were sources of plant occurrence data for the Pladias database. For information on the number of records and taxa integrated in the Pladias database, see Table 1. Descriptions are listed in the same order as in Table 1.

Database of the Distribution of Vascular Plants in the Czech Republic (FLDOK, Floristická Dokumentace)

The Database of the Distribution of Vascular Plants in the Czech Republic was established at the Institute of Botany of the Czechoslovak Academy of Sciences (now the Czech Academy of Sciences) by Pavel Tomšovic and Jan Štěpánek in 1992. It replaced hardcopy lists of floristic records compiled at the Institute since the early 1960s with the aim to facilitate the compilation of the Flora of the Czech Republic and species distribution mapping. The database management system was developed by Eduard Brabec in Paradox for DOS as a relational database, storing data in DBF files. The database was administrated in DOS graphical interface. The database includes mainly published records from the current Czech Republic. The records were extracted from national and local botanical and natural history journals. Most of the records are from the last few decades, but there are also many earlier records, some of them dating back to the early 19th century or even the late 18th century. Most of these old records were extracted from local floristic reports or studies on the distribution of particular species in this country. The records are localized in standard CEBA mapping grid cells (Niklfeld 1999) or their quadrants. More recent records are also localized by geographic coordinates. Most of the records include also a description of the locality, date and a reference to the published source. The taxa are recorded under the scientific names indicated in the original publication. No systematic quality check is applied, though some erroneous records are occasionally corrected based on feedback from users. In total 4,072 original sources were computerised till 31 December 2016, when the work on data collation ended. All data from FLDOK were transferred to the Pladias database and the original FLDOK is no longer being developed or filled with data.

Species Occurrence Database (NDOP, Nálezová databáze ochrany přírody)

The Species Occurrence Database was established at the Nature Conservation Agency of the Czech Republic (NCA CR) in 2006 by integrating previous database projects run by the nature conservation authorities. The database was built as a general database of biological records (also of animals and other organisms) for the Czech Republic. There are no taxonomic or temporal limits to the data included, but about 80% of the records were collected after the year 2000, and 75% of those were for vascular plants. One of the main database sources are the results of species and habitat mapping and surveys organized by NCA CR, which are included in the database on a compulsory basis. Other data sources are published records, imports of various experts' databases, imports of records from citizen science projects and data exchange with partners, such as with the CNPD database. The database has two main on-line interfaces: the editing application and the viewing application, which provide an effective way of data handling. Each record is georeferenced with the precision of the original source, and it also includes the date, author's name and data source as basic information. The records could be accompanied by several additional parameter values or notes on taxonomy, vegetation, collection circumstances etc. according to the methodology in a particular data source. Each record can be subject to validation (i.e. checking and correction) by NCA experts; the records of species protected by law are given priority in validation. The records of vascular plants were provided to the Pladias database on 6 February 2016. On 3 August 2018 NDOP comprised 23,680,400 records, of which 17,771,946 were for vascular plants. Most of these records (ca 11.3 mil.) come from the national habitat mapping project (Härtel et al. 2009). The rest are from other small projects and data integrated from other databases (mainly CNPD and DCFCS). In this paper, we report as NDOP only those data that were not taken from the other databases considered here. The original records are available through the viewing application (portal.nature.cz/nd) to registered users. The accredited experts and nature conservation bodies are eligible to view the entire database, incl. records of protected species, which are not accessible to the public. Grid maps of species distribution are available to the general public without registration (portal.nature.cz/kartydruhu).

Czech National Phytosociological Database (CNPD, Česká národní fytocenologická databáze)

The Czech National Phytosociological Database was established in 1996 at the Department of Systematic Botany and Geobotany (now Department of Botany and Zoology) of Masaryk University in Brno (Chytrý & Rafajová 2003, Chytrý & Michalcová 2012). The technical managers of the database were, successively, Milan Chytrý, Marie Rafajová, Ilona Knollová, Zdenka Preislerová, Štěpánka Králová, Dana Holubová (née Michalcová) and Ilona Knollová. The primary purpose of the database was to support the production of a modern vegetation monograph of this country. The database stores phytosociological relevés and other types of vegetation-plot records from the Czech Republic, with the oldest being from 1922. The relevés are extracted from monographs, scientific journals, manuscripts and research reports, including surveys of nature reserves, master and PhD theses; many relevés have been contributed by their authors without being published or included in a manuscript. Most of the relevés were recorded following the sampling methods of the Braun-Blanquet approach (Westhoff & van der Maarel 1973). Each relevé includes a list of vascular plants and in many cases also bryophytes and lichens from a plot, and their cover values on a semi-quantitative scale. These species lists are connected to "header data", which contain information on the locality, usually with geographic coordinates, author, date of recording and, if applicable, literature reference, assignment to vegetation type and other kinds of information. Systematic quality control is applied to various variables recorded in vegetation plots. The taxonomy and nomenclature of vascular plants are based on a checklist prepared by Prof. Harald Niklfeld and Dr. Walter Gutermann (both from the University of Vienna), which is an unpublished update of the Central European checklist compiled by Ehrendorfer (1973). The database is stored and managed in the Turboveg for Windows program (Hennekens & Schaminée 2001). Metadata about the database are available in the Global Index of Vegetation-Plot Databases (GIVD; Dengler et al. 2011) under the code EU-CZ-001. The database is a part of the international vegetation-plot databases European Vegetation Archive (EVA; Chytrý et al. 2016, www.euroveg.org/eva-database) and sPlot (Bruelheide et al. 2019, https://www.idiv.de/en/sdiv/working_groups/wg_pool/splot/splot_database.html). The database continues to be developed, which includes additions of new relevés and corrections of various types of mistakes based on feedback from users. Most of the data in the database are freely available for non-commercial use. All records of vascular plants from this database were transferred to the Pladias database, and updates added depending on the accumulation of new vegetation plots.

Flora of Eastern Bohemia (KVC, Květena východních Čech)

Collecting of plant distribution records for this database, coordinated by Zdeněk Kaplan (Institute of Botany, Czech Academy of Sciences) started in 2007 in support of the Red Data Book project of eastern Bohemia. The database includes mainly published records extracted from national and local botanical and natural history journals. Contributors are mainly field botanists conducting their research in eastern Bohemia. The records include a description of the locality, date of recording (if available), the name of its author and a reference to the published source. The records are also localized to the quadrants of the standard mapping grid and the phytogeographical (sub)district; many of them have geographic coordinates. The taxa are recorded under the scientific names indicated in the original publication, but each record is also assigned to the name adopted in Kubát et al. (2002). The records were originally entered in MS Excel files with a predefined structure and restricted range of values for several fields. These files are currently being transferred to the Pladias database.

Records from the Summer Schools of Field Botany of the Czech Botanical Society (FK CBS, Floristické kurzy České botanické společnosti)

Records from the botanical excursions held during the Summer Schools of Field Botany, organized annually by the Czech (formerly Czechoslovak) Botanical Society in various parts of the country, are collected, sorted and usually published as a supplement to the society's journal Zprávy České (formerly Československé) botanické společnosti. Currently there are about 590,000 records collected during 42 summer schools, which are partly stored in FLDOK (468,000 records from 32 summer schools) and partly in a separate dataset included within Pladias (123,000 from 10 summer schools).

Floristic Database of the South-Bohemian Branch of the Czech Botanical Society (JCP CBS, Floristická databáze Jihočeské pobočky České botanické společnosti)

The Floristic Database of the South-Bohemian Branch of the Czech Botanical Society was established in 2003. The software tool for collection and management of floristic records was developed by Vladimír Hans under the supervision of Martin Lepší, Petr Lepší and Milan Štech. This database includes records mainly from southern Bohemia with no time limit, but most are from the period after 1980. Many of them were published in local floristic reports, usually as results of floristic surveys, but there is also a considerable amount of unpublished data of members of the South-Bohemian Branch of the Czech Botanical Society. Radim Paulič, especially, provided many of his finds and records extracted from manuscripts of other botanists. Each record usually includes a description of the locality, date and a reference to the published source. All records are localized by geographic coordinates or by coordinates of the nearest village or town. The taxa are mainly recorded under the scientific names indicated in the original publication. However, in some cases, the names were converted to those adopted by Kubát et al. (2002). No systematic quality check is applied, although some erroneous records are occasionally corrected based on feedback from users. The database comprises more than 120,000 floristic records (see Table 1). All records included until 2014 are now also in the Pladias database. The database is still being supplemented with new records. All data are available for contributors, and limited information is available to the general public (http://jpcbs.prf.jcu.cz/vraticka/www). The database management system was originally developed in the MS Access environment. A new system allowing online database management was established in 2013 under the name Vratička. The database is located on the server of the Department of Botany, Faculty of Science, University of South Bohemia. This database system was also adopted by some other regional databases integrated into the Pladias database.

Floristic Database of the Moravian-Silesian Branch of the Czech Botanical Society (MSP CBS, Floristická databáze Moravskoslezské pobočky České botanické společnosti)

The Floristic Database of the Moravian-Silesian Branch of the Czech Botanical Society was established in 2015 by Petr Kocián and David Hlisnikovský after a pilot recording project (Kocián & Hlisnikovský 2014) and is administered by Petr Kocián. It uses the same database management system as the Floristic Database of the South-Bohemian Branch of the Czech Botanical Society. A part of the data is provided online to the public (www.nalezovka.cz). This database includes records mainly from the north-eastern part of the Czech Republic (northern Moravia and Silesia), but records from other parts of the country are also uploaded. Most records are unpublished field records supplied by members of the Moravian-Silesian Branch of the Czech Botanical Society collected since 2010, and in large part collected by David Hlisnikovský. Each record is georeferenced and includes at least a description of the locality, date and a reference to its source or author. The taxa are mainly recorded under the scientific names indicated in the original source or those from Kubát et al. (2002). No systematic quality check is applied, although some records are occasionally revised based on feedback from users. The database comprises more than 21,500 records and is continuously updated. All these records are now also in the Pladias database.

Floristic Database of the Vysočina Region (DKV, Databáze květeny Vysočiny)

The Floristic Database of the Vysočina Region was established in 2015 with support from the EEA and Norway Grants no. EHP-CZ02-OV-1-013-2014. The database was used to compile maps of rare species in the Vysočina region (see http://www.prirodavysociny.cz/cs/38/cevnate-rostliny). The input format of the floristic data was primarily adjusted to follow the format of the Pladias database. This database includes records mainly from the Vysočina region and comprises about 22,800 floristic records (June 2018). Geographic coordinates of the site or coordinates of the nearest village or town were added to all the records. About 43% of the records originated from recent botanical surveys of nature reserves by Libor Ekrt and Ester Ekrtová. About 30% of the records are finds of regionally rare and endangered plants extracted from the file cards deposited in the Museum Vysočiny Jihlava (Růžička et al. 1968–2011), which contain mainly finds of Ivan Růžička and records compiled from botanical papers. The rest of the data are recent records from personal collections particularly of Luděk Čech, Libor Ekrt and Ester Ekrtová and records from some earlier published sources. This database is still being supplemented with new records. All the data are available only within the Pladias database.

Recording Cards and Floristic Archive of Bohumil Slavík (Slavík's Cards, Slavík's Archive, Škrtací seznamy a Floristický archiv B. Slavíka)

This collection of recording cards was established at the Institute of Botany of the Czechoslovak Academy of Sciences for the development of a distribution atlas of vascular plants in the Czech Republic based on a grid template of 5 longitudinal minutes × 3 latitudinal minutes in the late 1960s (Slavík 1971). Originally, several volumes of the Distribution Atlas of Vascular Plants of the Czech Republic were planned, however, only four instalments were published with gradually decreasing number of collaborators (Slavík 1986, 1990, 1998, Štěpánková 2012). Three types of forms were used in the field, containing 187, 227 or 300 taxa; the latter were intended for use in the Carpathian part of the country. The field records were collected by B. Slavík, other employees of the Institute of Botany, professional botanists from other institutions and amateur botanists (altogether about 60 collaborators). Simultaneously, plant records were extracted from floristic studies, phytosociological relevés and diploma theses dealing with local floras; however, this recording ceased during the 1990s. Altogether, 5,002 forms were completed, of which 1,582 contained original field records. In 2010–2012 the forms with original records were computerized using a MySQL environment Janitor. After a preliminary taxonomic revision, the records were transferred to the Pladias database. The records from the remaining forms were not computerized since they contained published records already included in the FLDOK database, records from vegetation plots already included in the CNPD, records extracted from diploma theses, which are often contaminated by numerous misidentifications, as well as records from public outreach articles. The resulting dataset contained 96,163 distribution records from 702 quadrants (150 species per quadrant on average). In addition, local floristic lists, reports and other manuscripts provided by various people to B. Slavík were computerized as Floristic Archive of B. Slavík, currently containing 17,560 records, but many manuscripts from this archive have not yet been computerized.

The Database of the Czech Forest Classification System (DCFCS, Databáze lesnické typologie)

The Database of the Czech Forest Classification System consists of vegetation-plot records obtained during the forest site survey mainly in the 1950s to 1970s. The electronic database with original field protocols was established in the Forest Management Institute, branch office Brno, in 2000. Each plot record includes a basic description of the locality, geographic coordinates and dendrometric, soil and climate data. On 1 January 2014, the database contained records from 48,978 vegetation plots from the Czech Republic collected by 96 authors between 1926 and 2013. Data on vegetation are stored and managed in Turboveg for Windows (Hennekens & Schaminée 2001), environmental data in the MS Access and MS Excel (soil data). Metadata on this database are available in the Global Index of Vegetation-Plot Databases (GIVD; Dengler et al. 2011) under the code EU-CZ-002. All floristic data stored until 2014 were provided for the Pladias project.

Flora of the Ještěd Ridge (FJR, Flóra Ještědského hřbetu)

The mapping of the flora of the Ještěd Ridge (northern Bohemia) started in 1998 and lasted until 2004. In this period, presences of all the species of vascular plant were recorded systematically by P. Petřík in 192 grid cells of 1/256 of the basic field of the Central European floristic mapping (i.e. ca 0.52 km^2). About 40% of the species were mapped in more detail in quarters of these grid cells (ca. 0.13 km^2). These species were either diagnostic species of phytosociological units (mostly of alliances) or synanthropic, endangered or rare species. Crops and planted ornamental species were not recorded. In 2015, a dataset of 58,351 records (1,053 plant taxa) gathered within the mapping project (Petřík 2006b) was imported into the Pladias database.