Long Database Report

VegBank – a permanent, open-access archive for vegetation-plot data

Robert K. Peet, Michael T. Lee, Michael D. Jennings & Don Faber-Langendoen

Abstract: Rapid progress is being made in North American vegetation science through recent developments within the U.S. National Vegetation Classification (USNVC). Central to these advances are sharing, archiving, and disseminating field plot data, the fundamental data required for describing and understanding vegetation communities. VegBank (GIVD ID NA-US-002) is the vegetation plot database of the Panel on Vegetation Classification of the Ecological Society of America. VegBank is a stand-alone, Internet-accessible, vegetation plot archive designed to allow users to easily submit, search, view, annotate, cite, and download diverse types of vegetation data. The archive also includes embedded databases that contain classifications of vegetation and individual organisms, designed and implemented to track the many-to-many relationship betweens names and plant or community concepts, as well as alternative party perspectives on accepted taxa. The VegBank data model is also implemented in VegBranch, a desktop tool for data management and for uploading to and downloading from VegBank.

Keywords: classification; concept-based taxonomy; ecology; Ecological Society of America; Federal Geographic Data Committee.

Nomenclature: U.S.D.A. 2002 (http://plants.usda.gov; viewed February 2002).

Abbreviations: ESA = Ecological Society of America; NCEAS = National Center for Ecological Analysis and Synthesis; USFGDC = U.S. Federal Geographic Data Committee; USNVC = U.S. National Vegetation Classification.

Received: 23 August 2011 – Accepted: 30 November 2011 – Co-ordinating Editor: Florian Jansen.

Introduction

In the U.S. the description and classification of vegetation types has been unified under the U.S. National Vegetation Classification (USNVC; see USFGDC 2008, Jennings et al. 2009). The USNVC has been developed and is implemented as a joint effort among professional societies (the Ecological Society of America), non-governmental organizations (NatureServe), and governmental agencies, under the auspices of the Federal Geographic Data Committee (USFGDC) Vegetation Subcommitee (chaired by the U.S. Forest Service). The USNVC is dedicated to the quantitative description of vegetation types and their classification within a continuously evolving hierarchical classification system. The USNVC is predicated on open access and open input of field data, new or revised type descriptions, peer review, and permanent archiving and dissemination of information. The USNVC operates on the fundamental components of species taxonomy, growth forms, field plot data, and vegetation type descriptions. Information from these components flows along a prescribed pathway from raw data to the formal description and classification of vegetation types.

Although the absence of a unified classification to support vegetation science in the U.S. has recently been overcome by development of the USNVC, significant advances in describing U.S., and more broadly North American, vegetation have been limited by the number of observations needed to span the geographical and ecological diversity of the continent. Our response to this limitation has been establishment of VegBank, an open-access archive of field plot data developed and operated by the Ecological Society of America's Panel on Vegetation Classification.

VegBank was developed as a public vegetation plot repository. Such an archive is absolutely necessary for storage and preservation of valuable vegetation information, but its value would be severely limited without an easy access point for interested parties to view and download vegetation plots of interest. Furthermore, a publicly accessible archive is needed to provide documentation for vegetation descriptions in the literature, to support the evolving USNVC, and for replication and testing of newly proposed types.

Database construction

Most data archives dedicated to collecting and maintaining vegetation plot information have the luxury of specifying and limiting what sorts of data are considered for inclusion. VegBank, being a public archive, has minimal requirements for inclusion of plots, and thus must house many kinds of plot data. Plots may be from any time period, from any location (so long as the location is known), use almost any methodology, list plant taxa according to any robust taxonomic standard, and include or exclude a myriad of environmental variables. To build a database that would support such a vast range of data types, we held many meetings with plant ecologists from around the world to see how they structured and thought about their data, and then designed a data structure that would meet all these needs. A series of design meetings resulted in a general, broadly applicable data model for plot records; core elements are shown in Figure 1 and described below. Commonly used standards for representing data were adopted from partner groups. Unique accession codes across every major unit of data allow easy communication between VegBank and partners, including our client database, VegBranch. In the end, this resulted in a fairly complex schema with 55 data tables that incorporate most approaches to recording vegetation plot data. VegBank design work began in 1999 and an operational web version was released in 2004.

GIVD Database ID: NA-US-002		**Last update:** 2012-07-10
VegBank		
Scope: Vegetation plots from anywhere in the world can be uploaded and searched, but the primary emphasis is on support of the US National Vegetation Classification.		
Status: completed and continuing	**Period:** 1971-2007	
Database manager(s): Robert Peet (peet@unc.edu); Michael Lee (michael.lee@unc.edu)		
Owner: Ecological Society of America (ESA)		
Web address: http://vegbank.org		
Availability: free online	**Online upload:** yes	**Online search:** yes
Database format(s): PostgreSQL	**Export format(s):** MS Access, CSV file, XML	
Publication: [NA]		
Plot type(s): normal plots; nested plots; time series	**Plot-size range:** 12-1000 m²	
Non-overlapping plots: 22,629	**Estimate of existing plots:** [NA]	**Completeness:** [NA]
Total plot observations: 22,629	**Number of sources:** 58	**Valid taxa:** 8,312
Countries: CA: 2.0%; US: 98.0%		
Forest: [NA] — **Non-forest:** [NA]		
Guilds: all vascular plants: 100%		
Environmental data: altitude: 94%; slope aspect: 84%; slope inclination: 87%; surface cover other than plants (open soil, litter, bare rock etc.): 24%; soil pH: 5%; other soil attributes: 22%		
Performance measure(s): presence/absence only: 0%; cover: 100%; measurements like diameter or height of trees: 9%		
Geographic localisation: GPS coordinates (precision 25 m or less): 3%; point coordinates less precise than GPS, up to 1 km: 8%; small grid (not coarser than 10 km): 1%; political units or only on a coarser scale (>10 km): 99%		
Sampling periods: 1970-1979: 4.0%; 1980-1989: 25.0%; 1990-1999: 48.0%; 2000-2009: 20.0%		
Information as of 2012-07-19; further details and future updates available from http://www.givd.info/ID/NA-US-002		

Data standards

Of paramount importance to an open database such as VegBank, where a large community of users are continuously archiving, querying, and retrieving information, is establishment of (a) standard attributes for plot records entities, (b) a standard descriptive vocabulary, (c) standard metadata, and (d) a standard data exchange schema. There are two basic types of field plot records in VegBank: classification plots and occurrence plots. Classification plots provide the data required for quantitative definition of vegetation types. Occurrence plots document a less rigorous observation of a plant assemblage at a known location, but are sufficient to document the occurrence of a particular association or alliance at a location. For each plot record, required fields are those minimally needed to serve as either classification or occurrence plots. Optimal fields are identified as fields that, while not required, reflect the best practices when recording plot data (see Jennings et al. 2009, Appendix B, for required and optimal fields for classification and occurrence plot records [http://esapubs.org/archive/mono/M079/006/appendix-B.htm]).

Standard information for recording plot data includes the identifier code assigned to the record by the plot's author, the method used for placement of the plot, the observation start and stop dates. One of the two major hurdles in developing the quantity of field plot records needed is populating the database with existing field observations, or legacy data. Another hurdle is regularizing electronic recording of new field plot observations among field workers to be consistent with the USNVC standards (Jennings et al. 2007). Because of the variable nature of legacy plot data, the accuracy of the observation dates is an included data field. This field indicates whether, for example, the date of observation is accurate within a week, month, or year, allowing future users to screen potential plot records accordingly. This indication of precision may be critical for application of the data to phenology, or for resampling vegetation to assess change.
Data standards that apply to plant taxa observed in the plot include at least one record per taxon, and include multiple records when taxa are observed in multiple strata. For classification plots a comprehensive list of taxa is required. For occurrence plots, only names of the dominant taxa are required. (Occurrence plots, because they do not contain full floristic composition, are not included in the summary statistics in this paper or in the summary information in GIVD.) The taxon reference authority (*secundum* in the sense of Berendsohn 1995; also see Berendsohn et al. 2003, Franz et al. 2008, Jansen & Dengler 2010) is required for all taxa. Possible abundance values per taxon include the percent of aerial cover per taxon per stratum as well as the overall cover of the taxon across all strata. Taxon inference area is required; this is the area in square meters used to estimate the cover of a given taxon. Taxon inference area is usually equal to taxon observation area, but may be larger or smaller for a specific taxon to accommodate special circumstances such as the use of subplots. The taxon basal area may be included as an option. This is the total basal area of woody stems in m² / ha for a given taxon, usually for those with a tree growth form. A taxon stem count may also be provided.

Information about the plot's location include latitude and longitude of the plot's center point using the World Geodetic System 1984 datum (WGS84) in decimal degrees, and description of any adjustments. If the type of coordinates originally recorded for the plot location is different from WGS84 in decimal degrees, these coordinates should be provided along with the datum or alternative geographic projection with units. An estimated accuracy of the location of the plot should also be provided as a plot origin having a 95% or greater probability of being within a given number of meters of the reported location. The total area of the plot in square meters is required for classification plots. If subplots are used, the total area value should include the subplots as well as the interstitial space between subplots. A data field for the estimated size of the overall stand of vegetation in which the plot occurs should be recorded. Other location-related information that should be recorded include the country, state/province or other subnational jurisdiction, and the county, township, parish, or similar local jurisdiction.

The VegBank plot data model

The most fundamental unit in VegBank is the plot module, which contains data directly related to the vegetation plot (Fig. 1). The core tables and their key elements are as follows.

1. VegBank takes the vegetation-plot concept and splits it into two parts: the plot and the observation (of the plot). A plot, from the VegBank perspective, is a static area, with variables that do not change, such as geo-location, elevation, slope, aspect, and area. An observation is made at a plot and may be repeated at another time in the future. Attributes included in observation are those that may change over time, such as date sampled, relative groundcover, methodology, and dynamic environmental variables such as soil attributes and hydrology. The plot-to-observation database relationship is one-to-many, though in practice there is often only one observation per plot.

2. The taxon observation is the recording of the presence of at least one individual described as a particular taxon in a particular observation of a plot. Each time such a taxon observation is made, the name applied to the taxon is recorded, as well as the reference that was used to determine and/or define what was meant by the taxon. This is the basis for the concept-based taxonomy (discussed below) that allows VegBank to distinguish records from different time periods, regions, or taxonomic references that may use the same name in different manners. Observation to taxon observation is a one-to-many relationship, with one record in taxon observation for each taxon recorded in an observation.

3. Attributes applying to a taxon observation, such as percent cover, biomass, basal area, as well as individual stems sizes, are included or linked into the taxon importance table, which points to the correct taxon observation record. There may be multiple taxon importance records reflecting importance values of a single taxon in different vertical strata. Taxon observation to taxon importance is a one-to-many relationship, with zero, one, or multiple importance values stored for a taxon observation.

4. VegBank lists each taxon once per plot, but because VegBank has a separate taxon interpretation table linked to the taxon observation table, multiple and subsequent taxon interpretations may be made by users of VegBank, including the original plot author, to update taxonomy, resolve ambiguity, or correct errors. The original notation plus all subsequent taxon interpretations are retained so that it is possible to see varying opinions about a taxon observation over time by different parties. Taxon observation to taxon interpretation is a one-to-many relationship.

5. An observation may be mapped onto community types through community interpretations. Because a party may classify a plot to more than one community type simultaneously as part of the same classification effort due to lack of clarity, there is a container element called community classification that contains metadata about the classification method.

There are several other key portions of the VegBank data model, such as the reference and party modules. These were modeled after the Ecological Metadata Language (EML; http://ecoinformatics.org/software/eml/) specification. For soils, we reserved space for standard attributes (e.g., pH, clay, silt, sand), but owing to differences in lab techniques causing ambiguity in meaning of an attribute, we left most soil attributes to the user-defined section of our data model. Cover methods, stratum methods, revisions, user-defined variables, and projects are handled in a fairly generic manner. Details about these can be found on the VegBank website (http://vegbank.org).

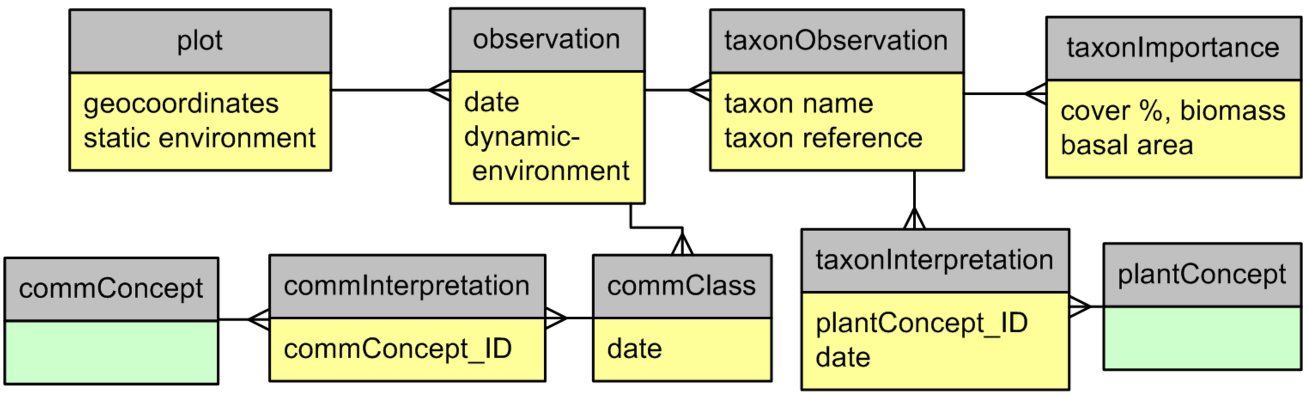

Fig. 1: The core VegBank data model.

Names
Carya ovata
Carya carolinae-septentrionalis
Carya ovata var. *ovata*
Carya ovata var. *australis*

References
Gleason 1952.
Radford et al. 1968.
Stone 1997.

Concepts (Name + Reference)
(All shagbark hickories)
C. ovata sec. Gleason 1952
C. ovata sec. Stone 1997

(Southern shagbark hickory)
C. carolinae-septentrionalis sec. Radford et al. 1968
C. ovata var. *australis sec.* Stone 1997

(Northern shagbark hickory)
C. ovata sec. Radford et al. 1968
C. ovata var. *ovata sec.* Stone 1997

Fig. 2: Taxon concepts for shagbark hickory. Four names are used in three references to indicate six taxonomic concepts which can be grouped to correspond to three meanings and either one or two sets of specimens.

The VegBank taxonomy data model

Classifications of organisms and communities provide a complex set of problems that must be addressed in any information system containing references to biological taxa or ecological communities. The core problem is that taxonomic standards vary with time, place, and investigator. Authoritative lists (e.g., Kartesz 1999, USDA PLANTS (http://plants.usda.gov), ITIS (http://www.itis.gov), NatureServe Explorer (http://www.natureserve.org/explorer/)) have been the traditional solution, but this approach fails to allow effective dataset integration for several reasons. (1) Online lists are periodically updated but usually are not simultaneously archived or archives of previous versions are not made available, with the consequence that the user cannot reconstruct the database as it was at some past time. (2) Ambiguity arises from the fact that a single name can be used for multiple taxonomic concepts and a single concept can be labeled with multiple names. (3) Different parties have different perspectives on acceptable names and their underlying concepts. The largest technical problem for VegBank to overcome was tracking plant and community names as they are treated in different manners across the different time periods and different regions, and by different people and organizations using different taxonomic standards.

A simple example that illustrates the taxon problem is provided by trees in the genus *Carya* referred to as shagbark hickory (Fig. 2). There are two different approaches taken in working with this particular group. One approach treats the group as a single taxonomic unit across its range, while a second recognizes one widespread taxon and another found exclusively in the south. The name *Carya ovata* as used in Gleason (1952) represents the first approach. The second approach, employed by Radford, Ahles, and Bell (1968) among others, recognizes *Carya ovata* as the widespread taxon and *Carya carolinae-septentrionalis* as the exclusively southern taxon. A third approach, advocated by Stone (1997), recognizes *Carya ovata* as a single species , with two varieties, *Carya ovata* var. *ovata* as the widespread taxon and *Carya ovata* var. *australis* as the southern taxon, which are identical in circumscription to the two Radford et al. species. If the name "*Carya ovata*" is encountered without qualification, it is not possible to know which of two possible concepts is intended: the larger one-taxon concept, or only the widespread taxon. Of course, there are examples that are much more complex and pose more intractable problems. Franz *et al*. (2008) present the details of *Andropogon virginicus* as recognized in Radford *et al*. (1968), but which when examined across a set of 8 treatments spanning 115 years shows that there are in fact 9 groups of specimens that have been assigned to 17 taxonomic concepts (composed of one or more of the 9 groups of specimens) and 27 scientific names. A similar solution is needed for vegetation classification, where the same name and authority may have different meanings, depending on the publication.

The solution to the ambiguity arising from the use of species or community type names alone is to use concept-based taxonomy, which combines a name with a reference that defines how the name is being used. The VegBank taxonomic data model adopts the idea of a concept by combining a name with a reference. In literature such concepts are typically identified with the format of name plus 'secundum' (or sec.) plus author plus date (e.g., *Carya ovata* sec. Gleason 1952) as proposed by Berendsohn (1995).

The VegBank data model is nearly identical for plant concepts and community concepts, though each set is implemented in its own set of tables. In each the Concept is the primary entity for labeling a community or an organism. This is the intersection of a Name and a Reference wherein the particular application of the name was defined. An Interpretation occurs when an organism or plot observation is labeled with a concept. As there can be many concepts that are near synonyms assigned during different determination events, one concept needs to be chosen as having a Status of standard for each taxonObservation. Although a name is a critical component of the definition of a concept, that name is not necessarily the name that some other party would choose to apply to the concept. A Usage is a party-specific application of a name to a concept, which allows names to change without changing the concepts. The model for community type determinations differs from that used for determinations of taxon observations in that multiple community concepts can be applied simultaneously in a classification event, each with a specified fit, thus better accommodating the continuous and somewhat stochastic variation of vegetation. The levels of fit conform to the standard 1 to 5 scale fuzzy

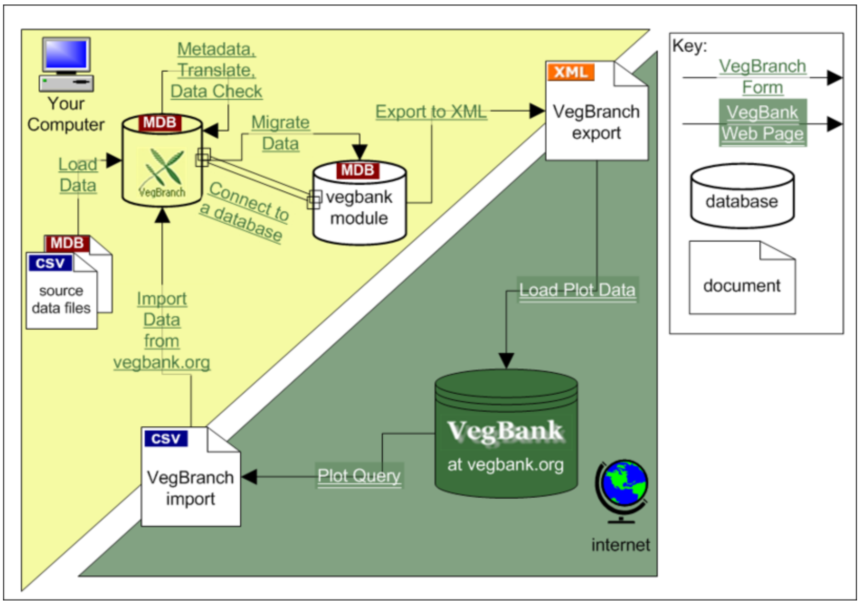

Fig. 3: The VegBank data system. The VegBank data archive can download data to the desktop VegBranch tool, which can in turn be used for local data management and to upload data to the VegBank archive

logic categories devised by Gopal and Woodcock (1994) of absolutely wrong, understandable but wrong, reasonable or acceptable, good, and absolutely correct.

The model we developed is similar to that proposed by the International Organization for Plant Information (IOPI; Berendsohn 1995, 1997, Berendsohn et al. 2003), although our model incorporates a number of advances over the IOPI model. For example, because the IOPI model is organized around fixed publication events, it does not efficiently map continuously changing party perspectives, making it difficult to provide software-generated, time-specific party views. The VegBank taxon concept model inspired creation of the TDWG Taxonomic Concept Schema (TCS; http://www.tdwg.org/standards/117/).

VegBank implementation

VegBank: a working archive

The VegBank database is deployed using PostgreSQL open-source relational database software. Standard international encoding UNICODE is used and data exchange is handled via the platform- and software-independent XML standard. We developed an automatically updatable XML schema (http://vegbank.org/xml) that can represent data in every part of the VegBank back-end database. Data model changes are automatically incorporated into the XML schema, as both are derived from the same source document that defines the model. Combined with accession codes, this enables easy and efficient communication of data between VegBank and the desktop client VegBranch (see below).

VegBank's underlying system is complemented by an intuitive user interface featuring the standard modern website components. User feedback has proven the design to be both pleasant and easy to use. Navigation is streamlined to allow one click access to every important page, and querying is built into every page. The logo and style create a look and feel that identifies VegBank. Presentation of all VegBank data is achieved through an interlinked system of JavaServer Pages (JSP). By applying a standardized development methodology, creation and maintenance of these views is greatly simplified. This approach ensures continuity of display as users navigate VegBank's data model. Certain views are also customizable so that users can select which attributes they prefer to see on the current page as well as pages they subsequently view.

Users can find data in several ways. Browsing pages allows users to simply walk through the data, which is essential for visually-oriented users who want to see the data to understand them. For searches, several levels of complexity are presented to allow both simple forms and complex criteria to pinpoint plots of interest. Once plots have been found, users may add plots to a "datacart," a personal collection of plots or other elements of interest. Datacart contents can be downloaded in a variety of formats. The default is a generic spreadsheet file, which is the fastest download; thousands of plots can be downloaded in a matter of seconds. Platform independent, full data model XML products can be downloaded, as can a file optimized for insertion into VegBranch. Users can also submit additional interpretations of plants found on plots and new interpretations of plot membership in communities. New plot data can be added from VegBranch or any other system capable of producing VegBank XML.

VegBank data are available for integration and viewing from other systems through REST-based web services, a standard URL request format. For example, NatureServe Explorer uses VegBank services to link to VegBank for all plots that belong in a specific community type. Citation of information in VegBank is easy because any VegBank entity that has an accession code can be viewed by a simple URL.

The VegBank data model and the embedded XML schema provided the initial model for a new international standard for exchange of vegetation plot data known as VegX (Wiser et al. 2011). We are looking forward to a revision of VegBank wherein VegBank XML is compliant with and an extension of VegX.

All components of the VegBank cyberinfrastructure are in the public domain and available for application and revision by any and all users. To obtain any part of the code base, simply visit https://code.ecoinformatics.org/code/vegbank/.

VegBranch: a tool for local management, preparation, and submission of plot data

To complement VegBank we designed and built a desktop client tool (VegBranch; Figure 3). Although the VegBank system is built almost exclusively around open-source software, VegBranch was built in MS Access to allow ease of user implementation and to allow average users to potentially "get-under-the-hood." VegBranch consists of a VegBank module that is an architectural mirror of the VegBank archive, and a VegBranch front-end that allows direct data import into a simplified table structure that is migrated to the VegBank module and subsequently exported to the online VegBank archive.

VegBranch allows direct data entry of simple-format data similar to that found in the NPS PLOTS database system (http://biology.usgs.gov/npsveg/tools/plot sdatabase.html) or the TURBOVEG database system (Hennekens and Schaminée 2001). Direct import from the NPS PLOTS database format is an available option. VegBranch is a powerful tool for migrating into VegBank diverse and complex legacy datasets. It prompts the user for missing metadata, and facilitates matching the user's data (e.g. plants and communities) to data already extant in VegBank. Error checking is quite thorough to prevent erroneous or partial datasets from being uploaded. VegBranch can also serve as a local database system for management of plot data. Plot data can be downloaded directly from VegBank to VegBranch and integrated with the local plot database to facilitate analyses.

Intellectual property

The most difficult non-technical challenge VegBank has to accommodate is protection of intellectual property. This is an issue in two ways. First, people naturally want to protect their data from being used before they are able to synthesize them and finish their initial publications. Secondly, plot records may contain sensitive location or endangered species information. Submitters of data wisely wish to protect the people, property, and plants from exposure that may cause harm.

VegBank addresses the first problem, that of restricting access to plots, with an embargo scheme that makes plots private until the person that submitted the plots lifts the embargo. This allows data to be deposited in VegBank while the data process is active and to leave a period of some months before the data are made public. The embargoed plots cannot be viewed in or downloaded from the system and are thus protected from early exposure to the public.

The second problem of protecting the exact location of endangered species has a number of possible solutions. A common solution employed in other database systems is simply to omit the endangered species from the list of taxa observed. This is an inadequate solution, because it fundamentally changes the composition of the plot, and the fact that the data are incomplete is either not known to the data user, or if it is made known, the identity of the rare species may be guessed based on habitat and the relatively small list of rare species that might need such protection. For these reasons, VegBank reduces the precision of the geo-coordinates of plots where the submitter needs to protect a species or landowner. The reduction in precision is elected by the plot submitter and may be to the nearest 0.01 degrees latitude and longitude (approximately 1 km), 0.1 degrees (approximately 10 km), or 1 degree (approximately 100 km). The location information may also be completely blocked, so that no geo-coordinates are reported, nor are any country, state, province, county locations given. Data such as directions to the plot are also blocked. This allows the plot composition to be reported unchanged (while keeping the plots broadly valid for analyses and vegetation type definition), but reduces the possibility of poaching endangered species. For users who require exact plot locations, other plots may serve their purposes, but for many applications, approximate location is suitable.

VegBank website

The VegBank website (http://vegbank.org) is free for use by the public. No registration is required, and anyone with an internet connection can browse or search through plots, plant concepts, community concepts, and any other public data in our system. Our searching mechanism relies on a compiled cache of keywords for the various types of data in our system, making it extremely powerful. VegBank does not require that you search for full words, so you can easily search for partial matches. The searching mechanism is both fast and accurate, and more than a few users have reported using VegBank to search through plants and communities, as the VegBank search is more accurate and efficient than the search functions on the original data source websites.

Once a user has discovered data of interest, particularly plots, these may be added to the user's datacart as described above. After compiling as large a list as the user would like, the data can be summarized into a constancy table, mapped using Google's mapping tools embedded into the VegBank website, and downloaded in several formats. Users that wish to save or build datasets over a longer period of time may register with the website and save multiple datasets according to their interests. Once a user has identified a plot or set of plots, it may be cited by using a unique accession code generated by VegBank. This provides a direct link to the full details of the plot or plots (see Plate A, B, and A). The following are examples of citations for a plot, plant concept, and community concept respectively:

- http://vegbank.org/cite/VB.Ob.3736.GRSM125
- http://vegbank.org/cite/VB.PC.54588.SPOROBOLUSAIROI
- http://vegbank.org/cite/VB.CC.5728.CEGL003906

Contributing data and opinions

Many users of VegBank have an interest not only in viewing plots, but also sharing their knowledge and perspective with other users of VegBank. These users may register and then request certification. This involves simply submission of a 1-page form describing the potential user's experience and expertise, which is then evaluated by a subcommittee of the ESA Panel on Vegetation Classification. Once certified, a user may annotate plots to correct, improve, or offer alternate opinions as to the correct placement of a plot in the community concept hierarchy, or correct concept to link to a taxon observation.

Certified users may also upload plots to VegBank. To upload plots or any other data to VegBank, users must create VegBank XML documents formatted according to our XML schema (http://vegbank.org/xml). As our data model is complex, and this schema allows for complete population of our data model, the XML schema is quite complex.

To assist users wishing to add their plots, we provide the previously described VegBranch tool, which walks the user through importing or entering plots, matching plant concepts to VegBank concepts, error-checking, and migration of data into a well-formatted VegBank XML document.

Users are strongly encouraged to enter their plots into VegBank, as this provides a permanent, accessible location for these data where they may then be viewed, reanalyzed, and appreciated by others. In this way the data provider can offload to VegBank time-consuming requests for data, and the burden of digital backups, not to mention the web security and setup that would be required for someone to host their own data. Journals may also be relieved of publishing the fine details of plots, as authors may simply cite the VegBank accession codes.

Plate: Vegetation types featured by the vegetation-plot database GIVD NA-US-002.

A Krummholz vegetation dominated by *Abies lasiocarpa* (sec. USDA PLANTS) at 3,500m in Rocky Mountain National Park, Colorado. VegBank plot 27228 (URL: http://vegbank.org/cite/VB.ob.27228.PEETROCKIES61) (Photo: R.K. Peet).

B Coastal marsh dominated by *Juncus roemerianus* and *Sabal palmetto*, Turtle Island, South Carolina. VegBank plot 26150 (URL: http://vegbank.org/cite/VB.Ob.26150.044040604) (Photo: R.K. Peet).

C Fire-maintained pine savanna dominated by *Pinus palustris*, and with *Ctenium aromaticum* in the foreground, in the Frances Marion National Forest, South Carolina. VegBank Plot 26057 (URL: http://vegbank.org/cite/VB.Ob.26056.027040459) (Photo: R.K. Peet).

Current VegBank data

At present there are already in VegBank more than 20,000 vegetation plots with the detail required to be used as classification plots in the sense of Jennings et al. 2008, plus more than 50,000 less detailed plots that mostly qualify as occurrence plots. These various plots contain approximately 8,000 species from projects around the United States and Canada. The primary data sources were USGS Gap Analysis Program plots, plot databases from the US National Park Service, plots from state natural heritage programs, and plots from the Carolina Vegetation Survey (Peet et al. 2012). Though VegBank's current set of plots are all from North America, plots from outside North America may be submitted by users. The diversity of vegetation types and plot locations represented in VegBank is illustrated in Plate A, B and C). There are currently 90,000 plant concepts, following United States Department of Agriculture (USDA) PLANTS database (2002 version; see http://plants.usda.gov) when possible, and 15,000 community concepts, following the USNVC when possible. Although confidentiality may be applied to a plot location or to block the plot from view until the author has published the initial results, the great majority of VegBank plots have no confidentiality restrictions. Additional information is available in the VegBank entry in GIVD (NA-US-002; see Dengler et al. 2011).

Vision for the future

In 2008 the U.S. Federal Geographic Data Committee (USFGDC) Vegetation Subcommittee adopted a new U.S. standard for the National Vegetation Classification (USNVC) system. Among other requirements, the USFGDC specifies that vegetation types in the USNVC have to be based on vegetation plot records available in a publicly accessible archive. While not specifically required, VegBank was the model for the proposed cyberinfrastructure and remains the only system that provides all of the mandated functionality. We anticipate that VegBank, or data systems based on its design, will become central to vegetation classification activities in the U.S. and elsewhere. Our goal is to encourage state, federal, academic and private programs that collect vegetation plot data to collaborate on data standards and move toward ready exchange of plot data. As a consequence, we anticipate that the availability of vegetation plots to support other forms of vegetation research will grow significantly and provide the basis for new and important directions of research.

Acknowledgements

VegBank is an activity of the Panel on Vegetation Classification of the Ecological Society of America (ESA). We gratefully acknowledge the support and encouragement of ESA, and especially the support of the ESA Science Office. The National Center for Ecological Analysis and Synthesis (NCEAS) in Santa Barbara, California provided start-up funding, hosted our meetings, and hosted our development staff. NCEAS staff members Matt Jones and Mark Schildhauer provided an enormous amount of critical advice and direction. National Science Foundation Grants DBI-9905838 and DBI-0213794 supported our development staff and meetings. The USFGDC Vegetation Subcommittee provided financial support for the ESA Vegetation Classification Panel and has consistently encouraged the ESA Panel to develop the information infrastructure necessary to support the USNVC. The U.S. Forest Service, the lead agency for the USFGDC Vegetation Standard, continues to provide maintenance funding. The U.S. Geological Survey provided support through its Gap Analysis Program and the National Biological Information Infrastructure. We thank NatureServe and its staff for their many and continued contributions to the USNVC.

Many individuals deserve special mention. The VegBank leadership team has at various times included the authors of this paper plus Dennis Grossman and Marilyn D. Walker. Our developers, John Harris, Gabriel Farrell, P. Mark Anderson, and Chad Berkley, were essential to the success of the project. Other important consultants during the design and testing of VegBank included Brad Boyle, Antoni Damman, Carol Fogelsong, Janet Gomon, Kathy Goodin, Stephan Hennekens, Paula Huddleston, Lee Ann Jacobs, Matt Jones, Larry Morse, Mike Palmer, Karen Patterson, Scott Peterson, Tom Philippi, Richard Pyle, Dave Roberts, Mark Schildhauer, Dave Tart, Steve Taswell, and Gary Waggoner.

References

Berendsohn, W.G. (1995): The concept of "potential taxa" in databases. – Taxon **44**: 207–212. CrossRef

Berendsohn, W.G. (1997): A taxonomic information model for botanical databases: the IOPI Model. – Taxon **46**: 283–309. CrossRef

Berendsohn, W.G., Döring, M., Geoffroy, M., Glück, K., Güntsch, A., Hahn, A., Kusber, W.-H., Li, J., Röpert, D., Specht, F. (2003): The Berlin model: a concept-based taxonomic information model. – Schriftenreihe für Vegetationskunde **39**: 15–42.

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

Franz, N. M., Peet, R.K., Weakley, A.S. (2008): On the use of taxonomic concepts in support of biodiversity research and taxonomy. – In: Wheeler, Q.D. [Ed.]: The new taxonomy. Systematics Association Special Volume **74**: 63–86. Boca Raton, FL: Taylor & Francis.

Gleason, H.A. (1952): The new Britton and Brown illustrated flora of the northeastern United States and adjacent Canada. – New York, NY: The New York Botanical Garden.

Gopal, S., Woodcock, C. (1994): Theory and methods for accuracy assessment of thematic maps using fuzzy sets. – Photogrammetric Engineering and Remote Sensing **60**: 181–188.

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

Jansen, F., Dengler, J. (2010): Plant names in vegetation databases - a neglected source of bias. – Journal of Vegetation Science **21**: 1179–1186. CrossRef

Jennings, M.D., Faber-Langendoen, D., Loucks, O.L., Peet, R.K., Roberts, D. (2009): Standards for Associations and Alliances of the U.S. National Vegetation Classification. – Ecological Monographs **79**: 173–199. CrossRef

Jennings, M., Faber-Langendoen, D.,Peet, R., Loucks, O., Glenn-Lewin, D., Damman, A., Barbour, M., Pfister, R., Grossman, D., Roberts, D., Tart, D. Walker, M., Talbot, S., Walker, J., Hartshorn, G., Waggoner, G., Abrams, M., Hill, A., Rejmanek, M. (2007): Description, documentation, and evaluation of associations and alliances within the U.S. National Vegetation

Classification. – Washington, DC.: Ecological Society of America, Panel on Vegetation Classification.
Kartesz, J.T. (1999): Synthesis of the North American flora. – North Carolina Botanical Garden, Chapel Hill, NC.
Peet, R.K., Lee, M.T., Boyle, M.F., Wentworth, T.R., Schafale, M.P., Weakley, A.S. (2012): Vegetation-plot database of the Carolina Vegetation Survey. – 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. (2012) [Eds.]: Vegetation databases for the 21st century. – Biodiversity & Ecology 4: 243–253. Hamburg. Biocentre Klein Flottbek and Botanical Garden. CrossRef
Radford, A.E., Ahles, H.E., Bell, C.R. (1968): Manual of the vascular flora of the Carolinas. – Chapel Hill, NC: University of North Carolina Press.
Stone, D.E. (1997): *Juglandaceae*. Flora North America. Volume 3. – Oxford, UK: Oxford University Press.
U.S. Federal Geographic Data Committee (2008): National Vegetation Classification Standard, Version 2. – FGDC-STD-005-2008 (Version 2). – URL: http://www.fgdc.gov/standards/projects/FGDC-standards-projects/vegetation/NVCS_V2_FINAL_2008-02.pdf.
Wiser, S.K., Spencer, N., De Cáceres, M., Kleikamp, M., Boyle, B., Peet R.K. (2011): Veg-X – an exchange standard for plot-based vegetation data. – Journal of Vegetation Science **22**: 598–609. CrossRef

Robert K. Peet* (peet@unc.edu) &
Michael T. Lee (michael.lee@unc.edu)
Department of Biology, University of North Carolina
Chapel Hill, NC 27599-3280, UNITED STATES

Michael D. Jennings (jennings@uidaho.edu)
Department of Geography, University of Idaho,
Moscow, ID 83844-3021, UNITED STATES

Don Faber-Langendoen (don_faber-langendoen@natureserve.org)
NatureServe
3467 Amber Road,
Syracuse, NY 13215, UNITED STATES

*Corresponding author