

News from the Global Index of Vegetation-Plot Databases (GIVD): the metadata platform, available data, and their properties

Florian Jansen, Falko Glöckler, Milan Chytrý, Miquel De Cáceres, Jörg Ewald, Manfred Finckh, Gabriela Lopez-Gonzalez, Jens Oldeland, Robert K. Peet, Joop H.J. Schaminée & Jürgen Dengler

Abstract: In 2010, we launched the Global Index of Vegetation-Plot Databases (GIVD; <http://www.givd.info>), an internet-based resource offering metadata of existing electronic vegetation databases. On 11 May 2012, 182 databases containing more than 2.8 million non-overlapping vegetation plots had been registered in GIVD. The majority of these plots were from European databases (123 databases, 1.87 million plots). The oldest plot record dated from 1864, but the vast majority of the plots had been collected since 1970. Most of the plots had areas between 1 and 1,000 m². A total of 68 databases also stored time series and/or nested-plot data. The vegetation-plot data registered in GIVD constitute a major resource for biodiversity research, not only through the large number of species occurrence records, but especially due to the storage of species co-occurrence information, complemented with site-specific structural data and plot-based environmental data. The increased ease of discovering and accessing such datasets, owing to their registration in GIVD, offers significant opportunities for large-scale studies in areas such as community ecology, macroecology, and global-change research. The results from such studies could be very important for nature conservation practice and policy.

Keywords: biodiversity; data sharing; ecoinformatics; metadata; phytosociology; relevé.

Abbreviations: GIVD = Global Index of Vegetation-Plot Databases.

Introduction

This report presents an update on the status of the Global Index of Vegetation-plot Databases (GIVD) published in *Journal of Vegetation Science* (Dengler et al. 2011). That article summarised the content of the databases registered in GIVD as of 30 December 2010. Since then, 51 new databases have been registered in GIVD and several other databases have refreshed the information about their content. Here we provide updated numbers and statistics of the metadatabase content as of 11 May 2012.

We defined vegetation-plot records, often called relevés, in a broad sense as records of plant taxon co-occurrences for a specified area, at a particular site, at a particular time. These records constitute the primary descriptive data on which much of vegetation science is based. Apart from the accurately defined place, area, and time of recording, a complete record of plant taxa (at least for a specified guild such as all trees) is crucial. As

can be seen in the present volume (Dengler et al. 2012), millions of vegetation plots have already been digitized, although mostly for local and regional purposes (also Schaminée et al. 2009). Discovery of, and access to, data in this massive and distributed resource has not only been facilitated by establishment of GIVD, but has also helped to change ways of working in this field so that more plot data are regularly becoming registered and available.

The *Global Index of Vegetation-Plot Databases* (GIVD) was launched in August 2010 (<http://www.givd.info>) as a metadata facility. GIVD is hosted at the Institute of Botany and Landscape Ecology of the University of Greifswald and is endorsed by the German Working Group on Vegetation Databases (<http://net.phyd.floraweb.de>), the International Association for Vegetation Science (<http://iavs.org>) including its European Vegetation Survey (EVS; <http://www.evitalia.eu>) and Eco-Informatics (<http://cvs.bio.unc.edu/iavs-ecoinfo/>)

working groups. Apart from access to all uploaded metadata, the GIVD homepage provides functions to search for databases that meet specific requirements and analytical functions such as summary statistics.

In this paper, we provide statistics on the content of GIVD as of 11 May 2012. However, GIVD is an ongoing project and you will always find the newest database information as well as up-to-date statistics at <http://www.givd.info>.

New data and new features in GIVD

As of 11 May 2012, a total of 182 databases had been registered in GIVD, an increase of 39% since 30 December 2010 (Fig. 1).

The 51 new databases contain 192,728 vegetation plots from 43 countries. At the same time, the number of plots in already registered databases increased by 170,531, which shows the rapid rate of

increase of plot data documented on the interactive GIVD platform. Altogether 2,807,960 vegetation plots had been registered by 11 May 2012. The largest number of new databases comes from Italy ($n = 16$), which was barely represented in the first round. This Italian contribution is followed by seven databases from Russia and seven from Germany. Apart from 40 new European databases, five multi-continental databases and five from Asia have been registered. Most databases in the world contain between 1,000 and 3,000 vegetation plots (Fig. 2).

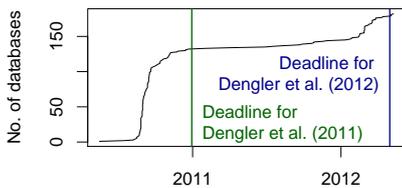


Fig. 1: Dates of database registrations. Most databases were registered shortly before the deadlines for reports in Dengler et al. (2011, *Journal of Vegetation Science*) and this volume of *Biodiversity & Ecology* (Dengler et al. 2012).

We asked all database curators to give additional information about habitat types represented in their databases when we sent them the proofs of their Fact Sheets published in this volume. The choices given are broad hierarchical “formations”, defined to give direct access to the main vegetation types. On the first hierarchy level (Fig. 3), we have made the distinction between forest (including woodlands) and non-forest vegetation, because this information can easily be derived for all plots; they mostly differ in sampled area, and they are usually analysed separately.

Terrestrial, aquatic or semi-aquatic (mires, swamps, etc.) habitats are distinguished on the second level. At the third level, alpine grasslands and arctic tundra are separated, whereas the non arctic-alpine vegetation is further divided into natural (savannas, deserts, coastal dunes, rocks), semi-natural (meadows, heathlands, garrigues) and anthropogenic vegetation (segetal and ruderal communities). In a future perspective, these broad formations can be extended by more detailed classifications, e.g. for Europe with the classes of the continental phytosociological classification (Rodwell et al. 2002, update by L. Mucina et al. in prep.) and for North America with the Macrogroups of the FDGC classification (Federal Geographic Data Committee 2008). As can be seen in Figure 4, forest and semi-natural vegetation are by far the most prominent vegetation types covered by vegetation databases, i.e. semi-natural vegetation is highly oversampled given its global frequency.

In addition to improvements in the information included in the metadatabase, some technical features of the web platform (<http://www.GIVD.info>) have been improved. Most notably, the search function has been extended to enhance usability. The first version provided only a simple search facility, enabling the user to type filtering keywords to select a list of available databases containing certain keywords in any of the metadatabase fields. Unfortunately, simple keywords often had no strong effect on filtering. This issue was solved by providing an optional extended search form, which enables the users to customize filter conditions dynamically. Five fields can be used to construct search filters: “Country ISO” (the 2-letter ISO country codes), “Country name” (English country name,

as well as its official name in the respective language), “GIVD ID” (the persistent identifier within GIVD, Dengler et al. 2011), “Database name” and “Contact name” (the names of all registered database curators and responsible researchers). These fields can be combined with the operators “AND” and “OR” to create diverse filter conditions. The usage of a placeholder (%) in the search phrase is also possible.

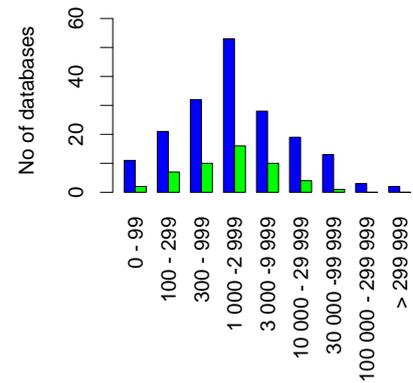


Fig. 2: Size distribution (in terms of the number of non-overlapping plots) of the 51 new databases (green) registered in GIVD between 30 December 2010 and 11 May 2012 compared to the overall size distribution (blue). Both barplots show a similar distribution with a mode at databases between 1,000 and 3,000 vegetation plots.

Another improvement is the direct resolution of the unique identifiers via GIVD’s web-domain. It is now possible to display the details of a particular database entry by using the web-address:

<http://www.givd.info/ID/<GIVD ID>>,

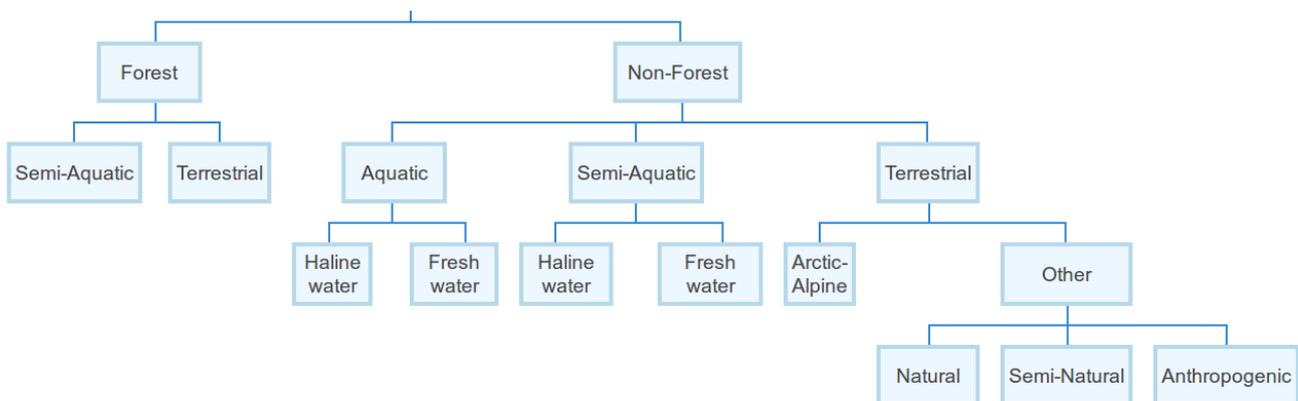


Fig. 3: Hierarchy of the broadly defined vegetation types given for the classification of GIVD-registered vegetation plots.

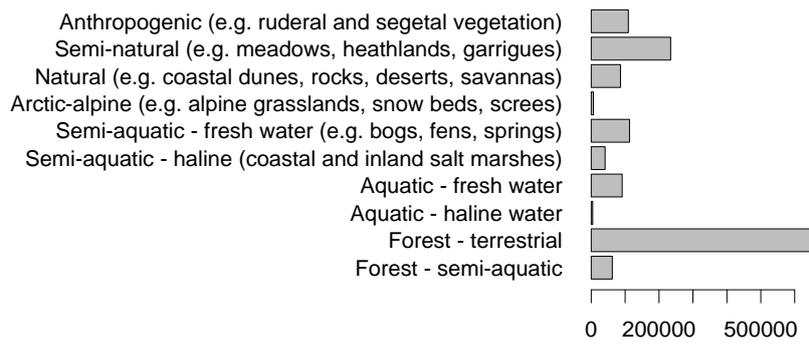


Fig. 4: Number of vegetation-plots within the vegetation types given in Figure 3 for those databases in GIVD for which a classification has been given. Most of the forest plots ($n = 538,428$) come from the North American forest inventory (NA-US-001).

where [GIVDID] represents the specific identifier (e.g. <http://www.givd.info/ID/EU-DE-001>). Because the database identifiers are unique in the context of GIVD, the combination with the web domain makes it a Universal Resource Identifier (URI). In this way, each meta-dataset can be directly linked from other web pages, thus enhancing usability and interconnections. As long as GIVD maintains the identifiers, the URIs can be treated as uniform persistent identifiers in the Internet (Berners-Lee 1994).

From the start of the GIVD project, it was the intention of the creators to provide not only an online metadatabase, but also an opportunity for the component datasets to be summarized and described in permanent, citable publications. Only

one third of all databases registered in GIVD as of 11 May 2012 mentioned a previous publication describing the content of the database (see the field “publication” in the Database Reports in this volume). Moreover, many of the extant descriptions had been published in regional journals that are often difficult to access. The present volume represents the culmination of this effort, and in addition provides a resource for not only discovering critical data, but also for examining the diversity of possible data models and data systems for managing vegetation-plot data. As an example, GIVD will provide a viable source for the information system *SynBioSys Europe*, an initiative of the European Vegetation Survey (Schaminée et al. 2007).

GIVD: a global biodiversity information platform

The overview in Dengler et al. (2011) showed a truly global distribution of information about available vegetation-plot data. As of 11 May 2012, GIVD contained information about vegetation plots from 121 countries, half of all existing countries in the world. In addition to the countries already covered by 30 December 2010, the new databases contain plot data from Bosnia and Herzegovina, India, Laos, Madagascar, Montenegro, Suriname, and Yemen. Compared to the global area of countries except Antarctica, GIVD contains vegetation-plot information about more than 80% of the continental area by countries (Fig. 5). Europe is the continent with the highest number of registered vegetation plots, and within Europe the Netherlands is the country with both the highest density and number of vegetation plots (Fig. 6), whereas Germany has the highest number of registered databases.

As of 11 May 2012, GIVD contained metadata for 182 vegetation-plot databases. Together, these databases contained 2.8 million non-overlapping vegetation plots (i.e. each time series and nested-plot series is counted only once), and 5.14 million vegetation-plot observations (which includes sub-plots in nested samples and repeated observations in time series). 68 databases (37%) contained at least some vegetation plots with repeated

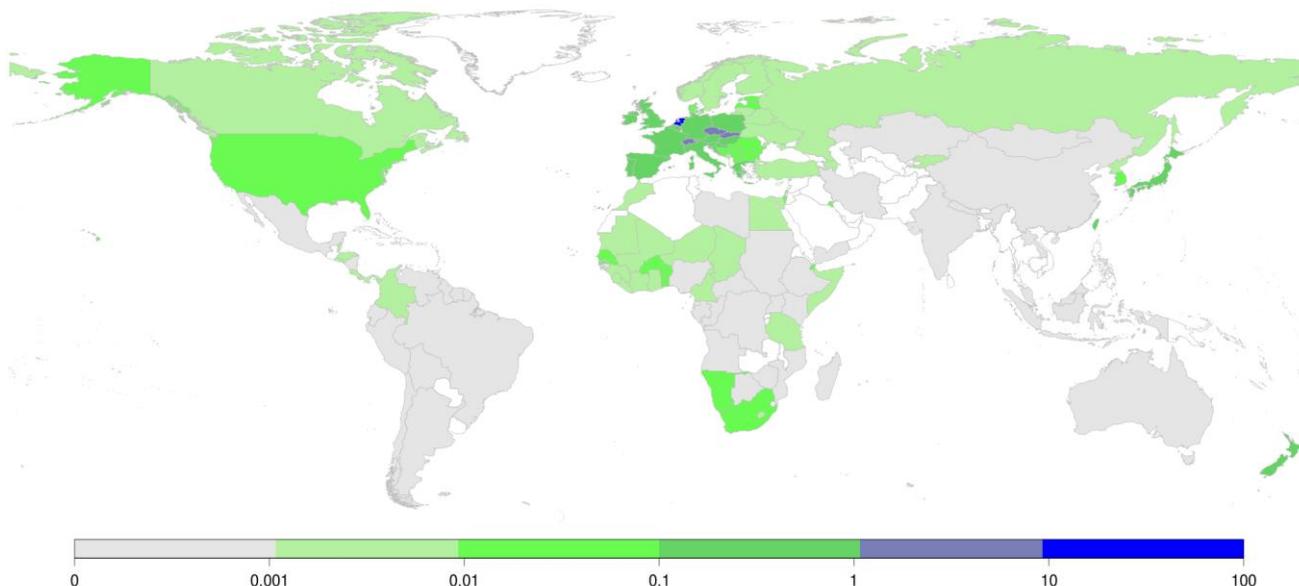


Fig. 5: Density of available, non-overlapping vegetation plots per km² for countries and equivalent geographical units of the world (based on GIVD as of 11 May 2012). None of the registered databases contained data for the countries in white. Note the non-linear density scaling. White colour means that no plot is registered from the respective territory, whereas a single plot from a country was sufficient for the country to be displayed in grey.

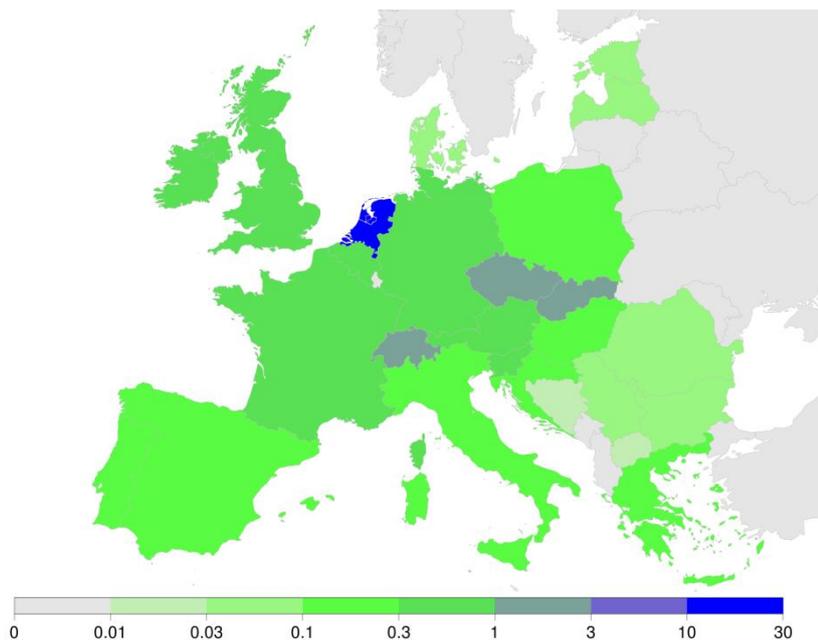


Fig. 6: Density of available, non-overlapping vegetation plots per km² for countries and equivalent geographical units in Europe (based on GIVD as of 11 May 2012). Note that the density scaling is non-linear and differs from Figure 5 in order to differentiate among the European territories.

observations and/or a nested design (sampling different plot sizes of the same area). For 79 databases with a total of 1,368,893 plots, i.e. 49% of all registered non-overlapping plots, database owners have mentioned that at least a part of the data is stored in TURBOVEG format (Hennekens & Schamineé 2007), which is therefore by far the most widespread vegetation database system worldwide.

Notwithstanding still missing databases and taking into account a small potential overlap in records, the number of non-overlapping plots in GIVD for Europe (1,868,457) now already exceeds the total number estimated by Schamineé et al. (2009: 1,852,000). Because of the long tradition of phytosociology, vegetation-plot databases present an astonishingly homogeneous data source with data collected using relatively consistent methodology. A major difference between databases is the distribution of plot sizes.

Figure 7 shows that there is a recent tendency to smaller plot sizes, especially plots less than 1 m² for more detailed vegetation analysis. Another difference is related to the sampled guilds (broad groups of species with similar resource use, e.g. trees, herbs, epiphytes) used to inventory the co-occurrence information. In contrast to Europe, in North America several databases are restricted to tree species without herb layer information. Throughout the world, information about

bryophytes, lichens, and algae in vegetation databases varies from scarce to absent (Fig. 8).

The comparison of the present distributions of vegetation-plot numbers per decade (Fig. 9) to that given in Dengler et al. (2011) shows that the new as well as the old, now updated, databases contain data from the current decade (since 2010), which obviously had to be largely missing in the last enquiry (Dengler et al. 2011). The high number already reported for the current decade suggests that we can expect researchers to collect, store and publish at least as many vegetation plots in this decade than in the last.

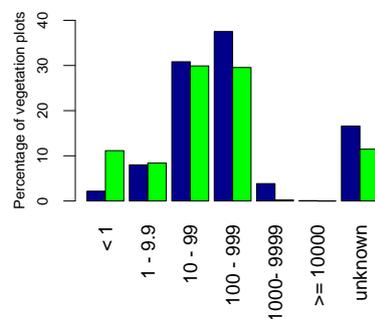


Fig. 7: Plot size distribution (in m²) of the newly registered vegetation-plots (51 databases registered between 30 December 2010 and 11 May 2012) in green colour, compared to that of the already existing databases (blue).

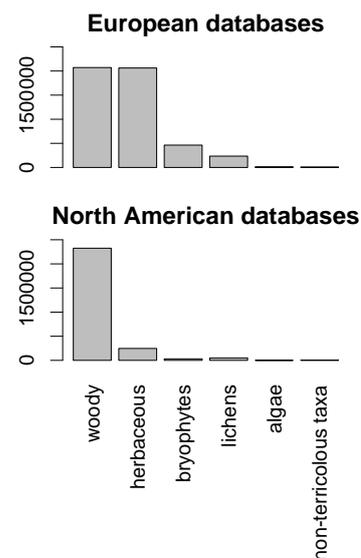


Fig. 8: Representation of (potentially) recorded guilds in 1,994,236 plot observations from Europe compared to the representation in 2,159,986 observations from North America registered in GIVD on 11 May 2012. Whereas in Europe nearly all databases include records of herbaceous plants (with a small number of exceptions like EU-CH-004 with only bryophytes or EU-CH-005, Swiss Forest Vegetation recording only trees), in North America several databases, but most prominently NA-US-001, the Forest Inventory National Database (FIADB), contain plot records with only records of tree species.

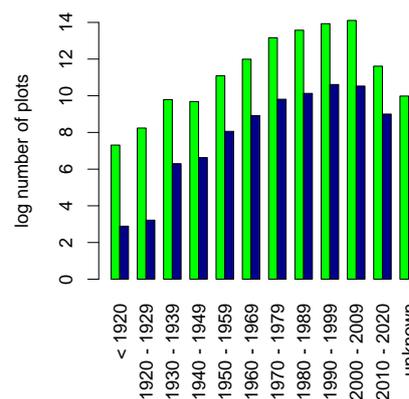


Fig. 9: Temporal distribution of the new vegetation-plot observations from 49 databases registered between 30 December 2010 and 11 May 2012 in green colour, compared to the temporal distribution of the new vegetation plots reported in updates for the previously registered databases. Note that the y-axis is the natural logarithm of the number of plots.

Web statistics of GIVD usage

We started logging usage data for the web domain *givd.info* in June 2011. The log files of the web-server have allowed us to analyse the number and duration of visits, the number of visitors, the kinds of web-browser used, the number and addresses of web-pages referring to GIVD, the amount of data transferred, as well as the most frequently used keywords in search engines looking for GIVD. Here we summarise the most important facts pertaining to users' activities during the 11-month period of June 2011 to April 2012.

We recorded 4,764 visitors (IP addresses) and 9,834 visits, whereas one visit is defined as a request to the website, with following requests only counted if the same visitor had been inactive for at least 60 minutes. On average, users visited GIVD twice from the same IP address.

How long the visitors 'stay' at our website can be estimated indirectly by looking at the differences between their first and their last request during their visit. 70% of the visitors stayed less than 30 seconds. This group mostly consisted of people who had only one request per visit (independently of how long they actually read the web-page), i.e. visitors who wanted to have a short view for different reasons as well as people who accidentally accessed the website. The other 30% of the visitors are almost equally distributed over the given categories for the duration of their visit: 30 s – 2 min, 2–5 min, 5–15 min, 15–30 min, 30 min – 1 h, > 1 h.

In general, there are three ways users can reach the GIVD website: (1) direct loading (e.g. by typing the URL into their browser's address bar or clicking onto a bookmark they saved before), (2) through a link provided on another website, or (3) by browsing with the help of a search engine. This behaviour has been recorded by our logs and is helpful for assessing the interconnection of *http://www.givd.info* within the Internet and the awareness of people of the domain name.

86.7% of all visitors came directly onto our website (i.e. the domain name seems to be easily recognizable), 7.7% via external links and 5.6% via search engines (Fig. 10). Between 2011 and 2012, there was a decrease in visits via external links (from 12% to 4.5%) and an increase in both direct access (83.5% to 89%) and access via search engines (4.4% to 6.5%).

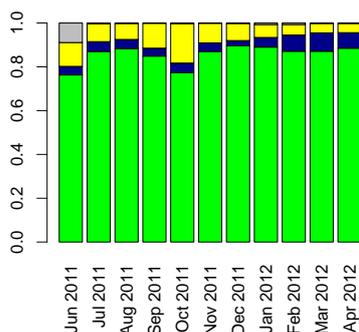


Fig. 10: Proportion of visitors to *http://www.givd.info* classified in respect to the referrer. The highest percentage (in green) came by directly entering the domain address or using bookmarks, an increasing proportion of visitors (blue) has found the portal by a search engine, a changing proportion was guided by an external webpage (yellow) or is unknown (grey).

The reasons for these changes can be very different, but we can assume that new users from 2011 clicked links in online publications and announcements, but as soon as they knew GIVD they accessed it directly. Since GIVD had more visits per month within the last year (Fig. 11), the searches of search engines contained more results linking to our website. However, the number of external links should be further increased in the future to make GIVD a fully interconnected web resource.

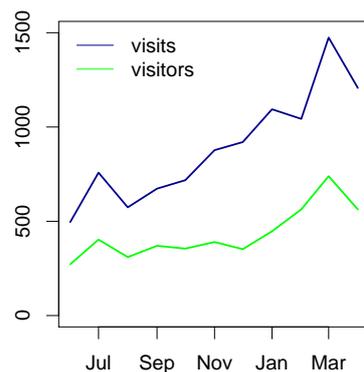


Fig. 11: Number of visitors and visits of *http://www.givd.info* from June 2011 until April 2012.

Acknowledgements

We are grateful to Will Simonson for polishing our English language usage.

References

- Berners-Lee, T. (1994): Universal Resource Identifiers in WWW: a unifying syntax for the expression of names and addresses of objects on the network as used in the World-Wide Web. – Request for Comments **1630**. URL: <http://wiki.tools.ietf.org/pdf/rfc1630.pdf>
- Dengler, J., Chytrý, M., Ewald, J. (2008): Phytosociology. – In Jørgensen, S.E., Fath, B.D. [Eds.]: Encyclopedia of ecology: 2767–2779. Oxford: Elsevier.
- Dengler, J., Jansen, F., Glöckler, F., Peet, R.K., De Cáceres, M., Chytrý, M., Ewald, J., Oldeland, J., Finckh, M., Lopez-Gonzalez, G., 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. [CrossRef](#)
- 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**. Hamburg: Biocentre Klein Flottbek and Botanical Garden. [CrossRef](#)
- Ewald, J. (2001): Der Beitrag pflanzensoziologischer Datenbanken zur vegetationsökologischen Forschung. – Berichte der Reinhold-Tüxen-Gesellschaft **13**: 53–69.
- Ewald, J. (2003): A critique for phytosociology. – Journal of Vegetation Science **14**: 291–296. [CrossRef](#)
- Federal Geographic Data Committee (2008). National Vegetation Classification Standard, FGDC-STD-005-2008 (Version 2). – Reston, Virginia: Vegetation Subcommittee, Federal Geographic Data Committee, FGDC Secretariat, U.S. Geological Survey.
- 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. [CrossRef](#)
- Mucina, L., Schaminée, J.H.J., Rodwell, J.S. (2000): Common data standards for recording relevés in field survey for vegetation classification. – Journal of Vegetation Science **11**: 769–772. [CrossRef](#)
- Rodwell, J.S. (1995): The European Vegetation Survey questionnaire: an overview of phytosociological data, vegetation survey programmes and data bases in Europe. – Annali di Botanica **53** 87–98.
- Rodwell, J.S., Schaminée, J.H.J., Mucina, L., Pignatti, S., Dring, J., Moss, D. (2002): The diversity of European vegetation – An overview of phytosociological alliances and their relationships to EUNIS habitats. – Rapport EC-LNV **2002(054)**. Wageningen: National

- Reference Centre for Agriculture, Nature and Fisheries.
- Schaminée, J.H.J., Janssen, J.A.M., Haveman, R., Hennekens, S.M., Heuvelink, G.B.M., Huiskes, H.P.J., Weeda, E.J. (2006): Schatten voor de natuur – Achtergronden, inventaris en toepassingen van de Landelijke Vegetatie Databank [in Dutch]. – Utrecht: KNNV Uitgeverij.
- Schaminée, J.H.J., Hennekens, S.M., Ozinga, W.A. (2007): Use of the ecological information system SynBioSys for the analysis of large databases. – *Journal of Vegetation Science* **18**: 463–470.
- Schaminée, J.H.J., Hennekens, S.M., Chytrý, M., Rodwell, J.S. (2009): Vegetation-plot data and databases in Europe: an overview. – *Preslia* **81**: 173–185.
- Thiers, B. (2010): Index Herbariorum: A global directory of public herbaria and associated staff. – New York: New York Botanical Garden.
- Florian Jansen* (jansen@uni-greifswald.de)
Landscape Ecology and Ecosystem Dynamics (LEED), Institute of Botany and Landscape Ecology, University of Greifswald
Grimmer Str. 88
17487 Greifswald, GERMANY
- Falko Glöckler (falko.gloeckler@gmx.de)
Biodiversity Informatics, Global Biodiversity Information Facility (GBIF), Natural History Museum Berlin
Invalidenstr. 43
10115 Berlin, GERMANY
- Milan Chytrý (chytry@sci.muni.cz)
Department of Botany and Zoology, Masaryk University
Kotlářská 2
611 37 Brno, CZECH REPUBLIC
- Miquel De Cáceres (miquelcaceres@gmail.com)
Centre Tecnològic Forestal de Catalunya
Ctra. St. Llorenç de Morunys km 2
25280 Solsona, SPAIN
- Jörg Ewald (joerg.ewald@hswt.de)
Botany & Vegetation Science, Faculty of Forestry, University of Applied Sciences Weihenstephan-Triesdorf
Hans-Carl-von-Carlowitz-Platz 3
85354 Freising, GERMANY
- Gabriela Lopez-Gonzalez (g.lopez-gonzalez@leeds.ac.uk)
School of Geography, University of Leeds
Leeds, LS2 9JT, UNITED KINGDOM
- Manfred Finckh (manfred.finckh@uni-hamburg.de), Jens Oldeland (jens.oldeland@uni-hamburg.de) & Jürgen Dengler (juergen.dengler@uni-hamburg.de)
Biodiversity, Evolution and Ecology of Plants (BEE), Biocentre Klein Flottbek and Botanical Garden, University of Hamburg
Ohnhorststr. 18
22609 Hamburg, GERMANY
- Robert K. Peet (peet@unc.edu)
Department of Biology CB#3280, University of North Carolina
Chapel Hill, NC
27599-3280 UNITED STATES

*Corresponding author