Sabine von Mering & Wolf-Henning Kusber

GBIF data network – infrastructure for biodiversity research
Open access to occurrence data of steppe species

Abstract

The Global Biodiversity Information Facility (GBIF) is an international initiative providing free and open access to biodiversity data via the Internet. Currently, more than 400 million records of all organism groups are available for scientific research, conservation, and education. GBIF-mobilised primary biodiversity data include records based on specimens from natural history collections and field observations as well as multimedia resources. Mobilisation and digitisation of herbarium data is one of the core activities of the botanical projects of GBIF Germany (www.gbif.de). The range of datasets is continuously extended by the inclusion of additional data from herbaria, from living collections (botanical gardens, *ex situ*-conservation), and from field observation projects (e.g. from floristic surveys and vegetation databases). GBIF Germany supports the scientific community by documentation and quality control of primary biodiversity data as well as hosting of data sources. Technologies developed in the context of BioCASe (Biological Collection Access Services) are used to publish annotations to digitised herbarium vouchers and to network data providers, e.g. within the *DNA Bank Network*. GBIF Germany will provide the image server technology for a joint *German Virtual Herbarium* and other services for the joint publication of multimedia data as well as metadata. Biodiversity data linked to accepted names and synonyms are used in taxonomic information systems, which are increasingly based on the *Internet Platform for Cybertaxonomy* software developed by the European Distributed Institute of Taxonomy (EDIT). GBIF provides an infrastructure supporting biodiversity research and conservation efforts by mobilisation, access, and analysis of information about the occurrence of various organisms over time and across the planet. GBIF Germany invites data holders from the steppe research community to share their data within the global GBIF data network.

Das GBIF Daten-Netzwerk – Infrastruktur für die Biodiversitätsforschung mit freiem Zugang zu Verbreitungsdaten von Steppenarten

Zusammenfassung

Die Global Biodiversity Information Facility (GBIF) ist eine internationale Initiative zur Förderung von freiem und kostenlosem Zugang zu Biodiversitätsdaten über das Internet. Aktuell sind mehr als 400 Mio. Datensätze aller Organismengruppen für Forschung, Naturschutz und Bildung verfügbar. Über GBIF zugängliche primäre Biodiversitätsdaten schließen sowohl Belegdaten aus naturhistorischen Sammlungen und Beobachtungsdaten als auch Multimediadaten ein. Zu den Kernaufgaben der botanischen Projekte innerhalb von GBIF Deutschland (www.gbif.de) zählen die Mobilisierung und Digitalisierung von Herbardaten. Durch die Aufnahme weiterer Datensätze aus Herbarien und Lebendsammlungen (Botanische Gärten, *ex situ*-Erhaltungskulturen) sowie Beobachtungsdaten aus Kartierungsprojekten und Vegetationsdatenbanken wird die Bandbreite verfügbarer Datensätze kontinuierlich erweitert. GBIF Deutschland unterstützt die wissenschaftliche Forschung durch Dokumentation und Qualitätskontrolle von Biodiversitätsdaten sowie das Hosting von Datenquellen. Sowohl bei der Publikation von Annotationen für digitalisierte Herbarbelege als auch beim Aufbau von Netzwerken verschiedener Datenanbieter kommt die BioCASe-Technologie (Biological Collection Access Services) zum Einsatz, z. B. innerhalb des *DNA-Bank-Netzwerkes*. GBIF Deutschland stellt die Bildserver-Technologie für ein gemeinsames *Virtuelles Herbarium Deutschland* und andere Services zur Publikation von Multimedia- und Metadaten zur Verfügung. Biodiversitätsdaten verknüpft mit akzeptierten wissenschaftlichen Namen sowie Synonymen werden in taxonomischen Informationssystemen verwendet, die verstärkt die im Rahmen des European Distributed Institute of Taxonomy (EDIT) entwickelte Software *Internet Platform for Cybertaxonomy* nutzen. Die GBIF-Infrastruktur fördert die Biodiversitätsforschung und den Naturschutz durch Mobilisierung, Zugang und Auswertung von Informationen zur Verbreitung verschiedener Organismen in Raum und Zeit. Datenanbieter aus der Steppenlebensraum-Forschung werden hiermit von GBIF Deutschland aufgerufen, weitere Daten zum globalen Daten-Netzwerk beizutragen.

1 Introduction

The Global Biodiversity Information Facility (GBIF, www.gbif.org) is an international initiative focussing on providing free and open access to biodiversity data via the Internet. Currently, more than 400 million records of all organism groups are available through the GBIF data portal (GBIF 2013, http://data.gbif.org, Fig. 1) and can be used for scientific research, conservation, and education. GBIF-mobilised primary biodiversity data include records based on specimens from natural history collections and field observations (i.e. which organism was observed where, when, and by whom) and linked multimedia resources.

Fig. 1:
Query on http://data.gbif.org for "scientific name" = *Stipa capillata* L. + "country" = Germany + "basis or record" = specimen. Name checked against Catalogue of Life, 32 specimens from collections worldwide, one specimen with image.

2 Botanical activities in GBIF Germany (GBIF-D)

Mobilisation and digitisation of herbarium data is one of the core activities of the botanical projects of GBIF-D. The range of datasets is continuously extended by the inclusion of additional data from herbaria, living collections (botanical gardens, *ex situ*-conservation), and field observation projects. Data from floristic surveys (Fig. 2) and vegetation databases are being included in GBIF (e.g. Ewald et al. 2012, Schaminée et al. 2012). Metadata of these data sources and others not yet connected to GBIF (e.g. Yamalov et al. 2012) are catalogued by Jansen et al. (2012).

Fig. 2:
Floristic maps of *Stipa capillata* L. for Germany. Source: Database FlorKart at the Bundesamt für Naturschutz, corrected as of 12/2006, see also Benkert et al. 1996, Haeupler & Schönfelder 1989. Published online as map by BfN (2012), data available via GBIF (GBIF 2013).

Table 1 gives an overview of quality aspects of biodiversity data relevant to vegetation science gathered in the GBIF Network. Primary biodiversity data in botany and vegetation science are generated with different scopes resulting in heterogeneous accuracy and precision of these data (see Chapman 2005).

Georeferenced occurrence data are most frequently requested from the GBIF portal (GBIF 2013) because they are visualised in maps on different scales (Fig. 3).

Dates, names of collectors, and persons who identified taxa complement primary occurrence data which are additionally accompanied by metadata describing source, scope and character of the dataset. Richness, correctness and accuracy of data from different sources (as indicated in Table 1) may help users to better understand information on single records provided by the GBIF network.

Table 1:
Overview of floristic and vegetation analyses and their representation in the GBIF Network.

	Floristic surveys and field mapping (Floristische Kartierung)	**Relevé (R), permanent plot (PP), transect (T)**	**Observation (O), specimen (S)**
Description	occurrence of a taxon in a specified area, e.g. TK 25 map	occurrence of taxa in a study site	single observation in the field (O) or single collection of a taxon in a collection (S)
Taxa included	1-∞	(1-)∞	1
Georeference	bounding box or Lat./ Long. plus uncertainty	bounding box or Lat./ Long. plus uncertainty	Lat./Long. plus uncertainty
Date	time span	single day (R, T), regular pattern of dates (PP)	normally single day or time span
Collected by	(1-)∞	1-∞	1
Identified by	(1-)∞	1-∞	1 (det., conf., rev.)
Data analyses (GBIF Portal)	all observations with the same time span and the same georeference pattern/same uncertainty	all observations with same date, georeference (R), same geo reference but different date (PP), same date and linear pattern of geo references (T)	
GBIF requirements	metadata, georeference and uncertainty as criteria to differentiate from single observations	metadata, identifier, collector, date, georeference and uncertainty as criteria to differentiate from single observations	metadata, identifier, collector, date, georeference and uncertainty, geo reference (O, S)
Best practise	see above (requirements)	image of the habitat (optional in BioCASe portals)	image of the observed taxon (O), of the specimen (S), of the habitat (optional in BioCASe portals)
GBIF record	single dot for a taxon within a map / quadrant	single observation within a relevé, plot, part of transect	single specimen (S) or observation (O)
GBIF example	Florkart	VegetWeb	herbaria (S), citizen science portals (O)

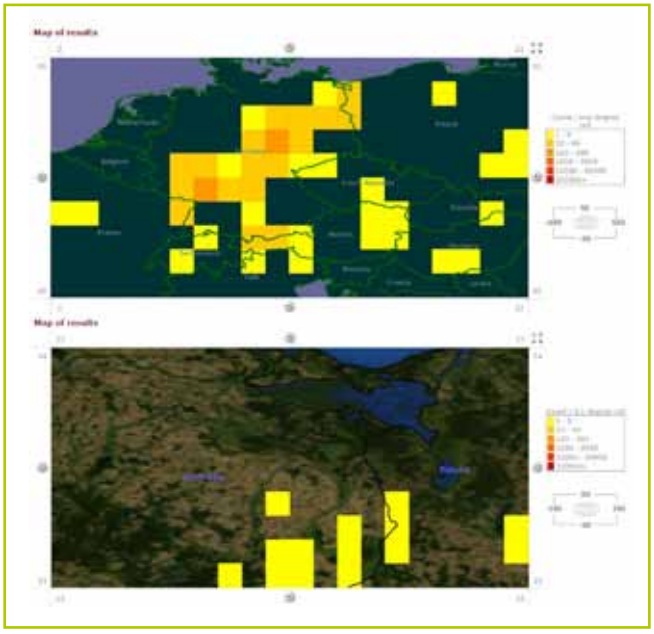

Fig. 3:
Maps of *Stipa capillata* L. for Central Europe (upper part, square 1° x 1°) and for the northernmost part of the German area (bottom, square 0.1° x 0.1°). Source: GBIF Data Portal http://data.gbif.org.

3 BioCASe as a standard Provider Software in Germany

GBIF-D supports the scientific community by documentation and quality control of primary biodiversity data as well as hosting of data sources. Technologies developed in the context of BioCASe (Biological Collection Access Services, www.biocase.org) are used to publish annotations to digitised herbarium vouchers and to network data providers, e.g. within the *DNA Bank Network* (www.dnabank-network.org, Droege et al. 2009, Gemeinholzer et al. 2010).

4 Portals for biodiversity research

Availability of multimedia objects (e.g. specimen images, digital photographs of plants, Fig. 4) is continuously increasing. GBIF-D is providing the image server technology and other services for the joint publication of multimedia data as well as metadata. High resolution image data are presented using the Image Server at BGBM. The FSI Viewer allows image enlargements comparable to a magnifying lens used in botanical field work (Fig. 4B-C). Plant details can be measured on the screen, facilitating a virtual herbarium visit. A *German Virtual Herbarium*, presenting specimen information and images of German herbaria, is in preparation.

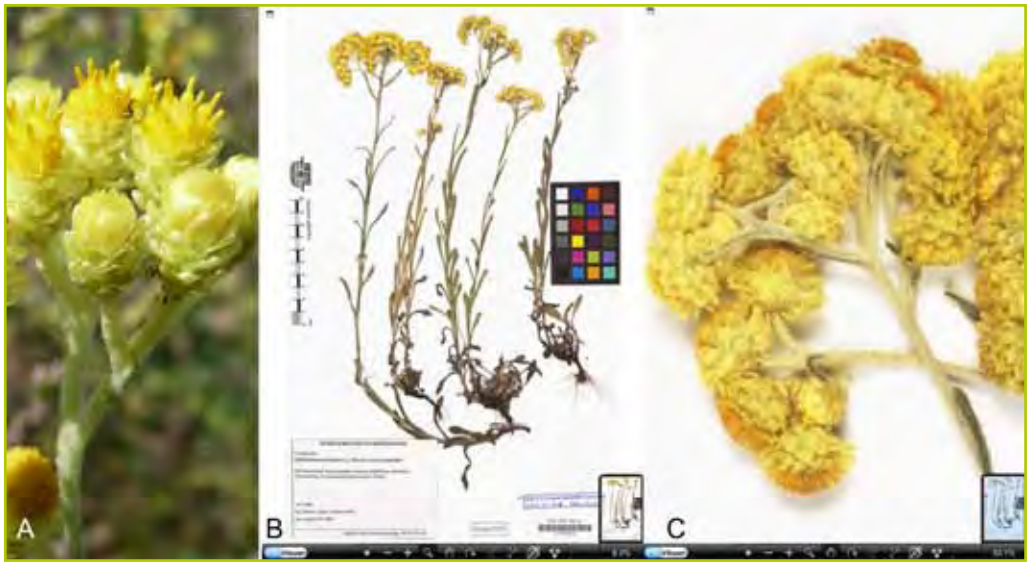

Fig. 4:
Helichrysum arenaria L.
A. Image documenting a field observation.
B – C. Display of specimen (Röpert 2000) and enlarged flower heads (C) using the FSI image viewer (BGBM, CC BY NC 3.0).

Biodiversity data linked to accepted names and synonyms are used in taxonomic information systems which are increasingly based on the *Internet Platform for Cybertaxonomy* software developed by the European Distributed Institute of Taxonomy (EDIT, www.e-taxonomy.eu, Berendsohn 2010, Berendsohn et al. 2011, Dengler et al. 2012). Examples include the Cichorieae Portal (http://wp6-cichorieae.e-taxonomy.eu/portal/) and checklists such as Euro+Med PlantBase (Euro+Med 2006-).

5 Summary and Outlook

GBIF provides an infrastructure supporting biodiversity research and conservation efforts by the mobilisation of and access to as well as analysis of information about the occurrence of various organisms over time and across the planet. As more data and also multimedia objects become available the usability is growing. Data quality is secured and improved by the development of new tools.
GBIF-D invites data holders from the steppe research community to share their data within the global GBIF data network.

Acknowledgements

The joint research project GBIF-D: "Kompetenzzentren innovativer Datenmobilisierung" is funded by the BMBF – Federal Ministry of Education and Research, grant: 01 LI 1001 A-F.

References

Benkert, D., Fukarek, F. & Korsch, H. (1996): Verbreitungsatlas der Farn- und Blütenpflanzen Ostdeutschlands. Jena.

Berendsohn, W. G. (2010): Devising the EDIT Platform for Cybertaxonomy. In: Nimis, P. L. & Vignes-Lebbe, R. (Eds.): Tools for identifying Biodiversity: Progress and Problems: 1–6.

–, Güntsch, A., Hoffmann, N., Kohlbecker, A., Luther, K. & Müller, A. (2011): Biodiversity information platforms: From standards to interoperability. In: Smith, V. & Penev, L. (Eds.): e-Infrastructures for data publishing in biodiversity science. ZooKeys **150**: 71–87, DOI: 10.3897/zookeys.150.2166.

[BfN] Bundesamt für Naturschutz (Ed.) (2012): FloraWeb. Published on the Internet http://www.floraweb.de [accessed 30.10.2012].

Chapman, A. D. (2005): Principles of Data Quality, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.

Dengler, J., Berendsohn, W. G., Bergmeier, E., Chytrý, M., Danihelka, J., Jansen, F., Kusber, W.-H., Landucci, F., Müller, A., Panfili, E., Schaminée, J. H. J., Venanzoni, R. & von Raab-Straube, E. (2012): The need for and the requirements of EuroSL, an electronic taxonomic reference list of all European plants. In: Dengler, J. et al. (Eds.): Vegetation databases for the 21st century. Biodiversity & Ecology 4: 15–24. DOI: 10.7809/b–e.00056.

Dröge, G., Zetzsche, H. & Gemeinholzer, B. (2009): Das DNA-Bank-Netzwerk – Eine Struktur für alle Fälle? In: Fischer, S., Maehle, E. & Reischuk, R. (Eds.): Informatik 2009 – Im Focus das Leben, Proceedings. Lecture Notes in Informatics, GI-Edition P-154: 26, 595–609.

Euro+Med (2006-): Euro+Med PlantBase – the information resource for Euro-Mediterranean plant diversity. Published on the Internet http://ww2.bgbm.org/EuroPlusMed/ [accessed 30.10.2012].

Ewald, J., May, R. & Kleikamp, M. (2012): VegetWeb – the national online-repository of vegetation plots from Germany. In: Dengler, J. et al. (Eds.): Vegetation databases for the 21st century. Biodiversity & Ecology **4**: 173–175. DOI:10.7809/b–e.00073.

[GBIF] Global Biodiversity Information Facility (2013). GBIF Data Portal. Version 1.2.6. Published on the Internet http://data.gbif.org [accessed 08.08.2013].

Gemeinholzer, B., Rey, I., Weising, K., Grundmann, M., Muellner, A. N., Zetzsche, H., Droege, G., Seberg, O., Petersen, G., Rawson, D. M. & Weigt, L. A. (2010): Organizing specimen and tissue preservation in the field for subsequent molecular analyses. In: ABC-Taxa, Volume 8 – Manual on Field Recording Techniques and Protocols for All Taxa Biodiversity Inventories, Chapter 7: 129–157.

Haeupler, H. & Schönfelder, P. (1989): Atlas der Farn- und Blütenpflanzen der Bundesrepublik Deutschland. 2nd ed., Stuttgart.

Jansen, F., Glöckler, F., Chytrý, M., De Cáceres, M., Ewald, J., Finckh, M., Lopez-Gonzalez, G., Oldeland, J., Peet, R. K., Schaminée, J. H. J. & Dengler, J. (2012): News from the Global Index of Vegetation-Plot Databases (GIVD): the metadata platform, available data, and their properties. In: Dengler, J. et al. (Eds.): Vegetation databases for the 21st century. Biodiversity & Ecology **4**: 77–82. DOI: 10.7809/b–e.0006

Röpert, D. (Ed.) (2000-, continuously updated): Digital specimen images at the Herbarium Berolinense. – Published on the Internet http://ww2.bgbm.org/herbarium/ (Barcode: B 10 0356769/ImageId: 316504) [accessed 30.10.2012].

Schamineé, J. H. J., Hennekens, S. M. & Ozinga, W. A. (2012): The Dutch National Vegetation Database. In: Dengler, J. et al. (Ed.): Vegetation databases for the 21st century. Biodiversity & Ecology **4**: 201–209.

Yamalov, S., Muldashev, A., Bayanov, A., Jirnova, T. & Solomesch, A. (2012): Database Meadows and Steppes of South Ural. In: Dengler, J. et al. (Eds.): Vegetation databases for the 21st century. Biodiversity & Ecology **4**: 291.

Addresses of authors

Sabine von Mering
Wolf-Henning Kusber
Botanic Garden and Botanical Museum Berlin-Dahlem
Freie Universität Berlin
Königin-Luise-Str. 6-8
14195 Berlin
GERMANY

E-Mail: s.vonmering@bgbm.org
E-Mail: w.h.kusber@bgbm.org