Open Access to Biodiversity Collection Data – GBIF Germany and the Botanical Node

Agnes Kirchhoff[1], Andrea Hahn[2], Jörg Holetschek[1], Patricia Kelbert[1,3], Regine Jahn[1] & Walter G. Berendsohn[1]

[1]Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Königin-Luise-Str. 6-8, 14195 Berlin, Germany; a.kirchhoff@bgbm.org; j.holetschek@bgbm.org; p.kelbert@bgbm.org; r.jahn@bgbm.org; w.berendsohn@bgbm.org
[2]GBIF Secretariat, Universitetsparken 15, 2100 Copenhagen, Denmark; ahahn@gbif.org
[3]Muséum National d'Histoire Naturelle, 45 rue Buffon, 75231 Paris Cedex 05, France

THE DATA

Over the last centuries, biologists have collected a huge amount of data about the world's species, and have stored and documented these primary biodiversity data in literature and as specimens and samples in natural history museums and living collections. These data are now being digitised and recorded in collection databases and as observation record sets (e.g. of floristic and faunistic survey and mapping projects as well as presence/absence data from ecological research plots).

GBIF, the Global Biodiversity Information Facility, provides a standardised global infrastructure to provide open access to primary biodiversity data over the Internet. Specialised Internet portals with enhanced search and retrieval functions provide access to different user communities. For instance, digitised biodiversity information of networked databases from all over the world is retrievable via the portal of the GBIF Secretariat in Copenhagen, while one of the two portals of the German Botanical Node focuses on facts about species and specimens of German flora by integrating the available specimen data with checklist information on the national level. This joint access to distributed databases facilitates new approaches for analysis and evaluation of formerly separate and inaccessible data, and will open a wide range of new research opportunities.

STRUCTURE OF GBIF INTERNATIONAL AND GBIF GERMANY

- GBIF International (GBIF 2007) to date (August 2007) comprises 40 countries and 33 organisations which have become GBIF participants by signing a Memorandum of Understanding (http://www.gbif.org/GBIF_org/participation). Participants agree to support network nodes to facilitate access to biodiversity data (MoU 2006).
- GBIF Germany (GBIF-D 2007a) is organised into several nodes, each responsible for a defined group of organisms, together covering the entire spectrum. The nodes are co-ordinated by institutions that are recognised authorities for their respective taxonomic group.
- GBIF-D Botany (GBIF-D 2007b), the German Botanical Node is co-ordinated at the Botanic Garden and Botanical Museum Berlin-Dahlem.

CONNECTING HISTORICAL DATA WITH PRESENT DAY MEASUREMENT DATA

Within the framework of the German GBIF node system, the Botanical Node provides access to botanical collections and databases in Germany and to data on the German flora. These data are contributed to GBIF International. Five information domains (Fig. 1) cover central aspects of botanical data provision. Some of them are already represented on the Internet with extensive information. GBIF-D Botany integrates these resources by combining them in a joint access system. As a result, biodiversity information held in herbaria, stored over a period of almost three centuries, can now be linked with present-day observation and collection data. Combining these data of multiple resources offers new information that would not exist in any single data source. A possible application could be the tracking of today's invasive species' progress back in time.

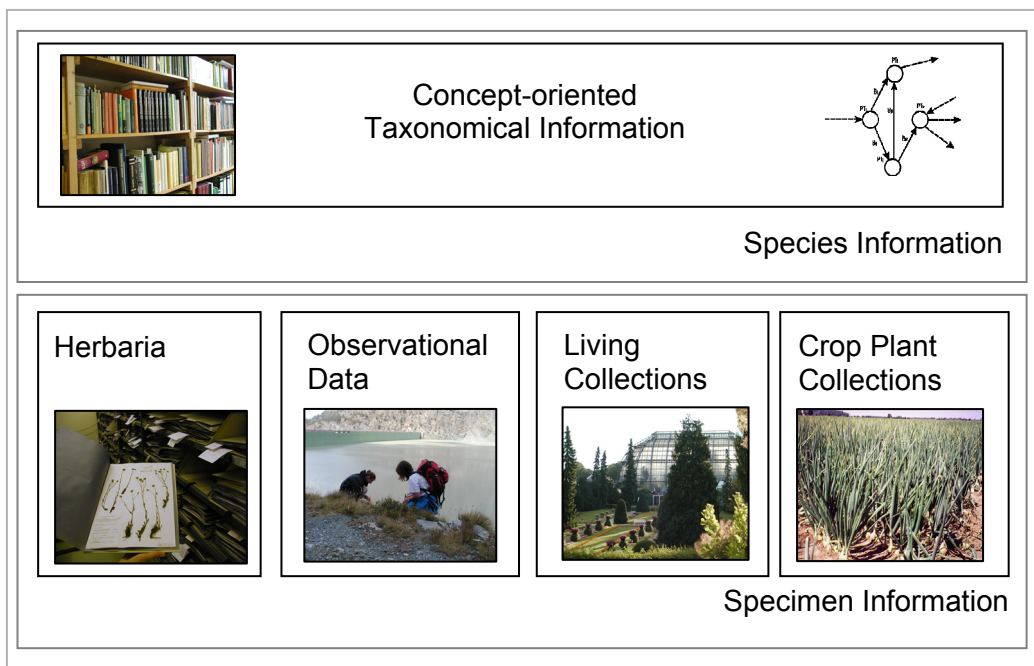### GBIF Botany information domains

Fig. 1. Information domains related to the Botanical Node of GBIF.

NETWORKING HETEROGENOUS BIODIVERSITY DATA – NETWORK ARCHITECTURE AND SOFTWARE COMPONENTS –

Historically, digital data storage of collections uses a wide variety of database management systems and software, and the same applies to data structures. With the help of software applications developed by publicly funded projects within the biodiversity research community, a search across distributed heterogeneous databases is now possible. One of these applications was developed by the EU-financed project BioCASE (Biological Collection Access Service for Europe, see BioCASE 2007); for a given search, a metadatabase is consulted to find relevant registered providers. The search is translated into a common query format (the BioCASe protocol), and is sent to all relevant providers via http. Each provider database is set up with a piece of software (the BioCASe provider software) that is responsible for translating the generic query into the native database language and structure; this setup has to be configured only once. The BioCASe provider software then translates the database query results into a common transfer format (XML structure), and returns it to the querying portal for further processing and display (Güntsch et al. 2007). The XML structure follows the ABCD (Access to Biological Collection Data) schema definition

(Berendsohn 2007), an XML standard defining more than 1000 terms describing specimen data. The ABCD standard was ratified by the organisation for Biodiversity Information Standards (TDWG) in 2006.

GBIF-D BOTANY DATA ACCESS

Starting in 2002, the Botanic Garden and Botanical Museum Berlin-Dahlem (BGBM) has set up the Botanical Node within the German GBIF. The node combines specimen data of various herbaria and collections of plants and algae. This is supplemented by data on today's flora and vegetation in Germany (FloraWeb 2007), supplied by the German Federal Agency for Nature Conservation (BfN), by data from the Systax information system on botanical gardens (Hoppe et al. 2007), and by the gene bank accession data of the Leibniz Institute of Plant Genetics and Crop Plant Research, IPK (GBIS/I). Information can be retrieved through two portals (www.gbif.de/botanik/datenabfrage) that provide access to distributed databases using BioCASE software.

a. GBIF-D Dataportal 'German Flora' Search

c. Display of a specimen picture

b. Specimen information

Fig. 2. *Triceratum favus* Ehrenb. – Biodiversity information of 1839 available via GBIF-D on the Internet in 2007.

One portal focuses on digitised botanical collections and databases held in German institutions (www.gbif.de/botanik/datenabfrage/default.html), while the other provides botanical records (at present more than 3.8 million) from collecting or observation localities within

Germany (http://search.biocase.org/gbif-de-botany). Running an advanced search on the latter portal allows the user to specify a number of search criteria; it is possible to search for a common name or higher taxon as well as a scientific name (Fig. 2). In addition, search results can be narrowed down to either observational or specimen data. To improve performance, the results are first shown in an overview table, which is linked to lists of detailed results. A supplemental download of the complete results is offered. Different data types, such as multimedia data (pictures, videos) as provided e.g. in the AlgaTerra Information System on Algae (Jahn & Kusber 2007, Kusber & Jahn 2007), are also available (Fig. 2).

At present the German Botanical Node provides to the international GBIF network 29 German Botanical collection databases (from 17 institutions) with more than 4.5 million records, which has helped GBIF Germany become the fifth largest data provider within GBIF International.

ACKNOWLEDGEMENTS

REFERENCES

Berendsohn, W. G. (ed.) 2007: The ABCD Schema - Access to Biological Collection Data. A joint CODATA and TDWG initiative supported by GBIF [online]. – Botanic Garden and Botanical Museum Berlin-Dahlem. [cited 2007-09-30]. Available from <http://www.bgbm.org/TDWG/CODATA/Schema/>.

BioCASE 2007: Biological Collection Access Service [online] BioCASE Secretary, Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin. [cited 2007-09-30]. Available from <http://www.biocase.org/>.

FloraWeb 2007: FloraWeb - Daten und Informationen zu Wildpflanzen und Vegetation Deutschlands [online]: Bundesamt für Naturschutz. [cited 2007-09-30]. Available from <http://www.floraweb.de/>.

GBIF 2007: GBIF Data Portal [online]. – Global Biodiversity Information Facilty [cited 2007-09-30]. Available from <http://www.gbif.org/>.

GBIF-D 2007a [online]: GBIF Deutschland. Informationen zum Aufbau des Deutschen Knotens der Global Biodiversity Information Facility [online]. – Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin. [cited 2007-09-30]. Available from <http://www.gbif.de/>.

GBIF-D 2007b: GBIF Deutschland Botanik [online]. – Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin. [cited 2007-09-30]. Available from <http://www.gbif.de/botanik/>.

GBIS/I – Genbankinformationssystem des IPK Gatersleben [online]. – Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK). Gatersleben. [cited 2007-09-30]. Available from <http://gbis.ipk-gatersleben.de/gbis_i/>.

Güntsch, A., Kusber, W.-H., Döring, M., Ciardelli, P. & Berendsohn, W. G. 2007: Common access to distributed biodiversity information. – Pp. 45-48 in: Kusber, W.-H. & Jahn, R. (ed.): Proceedings of the 1st Central-European Diatom Meeting 2007. – Berlin. [CrossRef]

Hoppe, J. R., Boos, E., Ludwig, T. Wiedemann, M. & Stützel, T. (ed.) 2007: SysTax - a Database System for Systematics and Taxonomy [online]. – Universität Ulm. [cited 2007-09-30]. Available from <http://www.biologie.uni-ulm.de/systax/index_d.html>.

Jahn, R. & Kusber, W.-H. (ed.) 2007: AlgaTerra Information System [online]. – Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin. [cited 2007-09-30]. Available from <http://www.algaterra.org>.

Kusber, W.-H. & Jahn, R. 2007: AlgaTerra Information System – Types data and data types. – Pp. 97-100 in: Kusber, W.-H. & Jahn, R. (ed.): Proceedings of the 1st Central-European Diatom Meeting 2007. – Berlin. [CrossRef]

MoU 2006: Memorandum of Understanding for the Global Biodiversity Information Facility [online]. – Global Biodiversity Information Facility. [cited 2007-09-30]. Available from <http://www.gbif.org/GBIF_org/ GBIF_Documents>.