

Beitr. Ent.	Keltern	ISSN 0005 - 805X
55 (2005) 2	S. 433 - 444	27.12.2005

Freier Zugang zu den Informationen der Artenvielfalt – Wie werde ich Teil der Global Biodiversity Information Facility (GBIF)?

Mit 5 Figuren

JÖRG HOLETSCHEK und WALTER G. BERENDSOHN

Zusammenfassung

Die Global Biodiversity Information Facility ist ein internationaler Zusammenschluss von Staaten und Organisationen mit dem Ziel, die weltweit vorhandenen Primärdaten zur Artenvielfalt über das Internet frei verfügbar zu machen. Dieser Artikel gibt einen kurzen Überblick über die Ziele und den Aufbau von GBIF sowie die zugrunde liegenden Prinzipien, erläutert Architektur und Datenströme im Netzwerk und geht auf die verwendeten Standards, Protokolle, Datenschemata und deren Implementierung ein. Im Abschluss werden die notwendigen Schritte zur Anbindung einer bestehenden Sammlungsdatenbank an das GBIF-Netzwerk aufgezeigt sowie die Möglichkeiten einer Hilfestellung durch GBIF-Deutschland.

Summary

The Global Biodiversity Information Facility is an international initiative of countries and organizations that aims at providing free access over the internet to the world's existing primary data on biodiversity. This article gives a short review of the goals and the organization of GBIF and its basic principles, illustrates the architecture and data flows within the network and illuminates the standards, protocols and schemas used for these purposes and their implementation. It describes the required steps for connecting an existing collection database to the GBIF network and the potential support by the GBIF helpdesk.

Key words

Biodiversity, GBIF, GBIF-D, database, network, collection, specimens, observations, XML, ABCD schema, provider software, Python wrapper, BioCASE protocol, mapping, search portal

Einleitung

Die Global Biodiversity Information Facility wurde im Jahre 2001 gegründet und hat sich zum Ziel gesetzt, die weltweit vorhandenen wissenschaftlichen Informationen zur Artenvielfalt in digitaler Form über das Internet frei und dauerhaft verfügbar zu machen. Das in Sammlungen und Forschungseinrichtungen vorhandene Wissen über die Biodiversität soll zu einem Netzwerk verknüpft und damit eine virtuellen Forschungseinrichtung geschaffen werden.

Unter Biodiversitätsdaten versteht man unter anderem Daten zum Vorkommen, zur Benennung, zur Beschreibung, zur Identifizierung, zum Schutz und zur nachhaltigen Nutzung der Organismen in unserer Umwelt (Oehlschlaeger & Berendsohn 2005). GBIF legt den Schwerpunkt in der Anfangsphase auf die Primärdaten zur Biodiversität, d.h. Informationen, die der direkten Beobachtung von Organismen in der Natur entspringen. Dazu zählen vor allem die 2-3 Milliarden von Objekten in wissenschaftlichen Forschungssammlungen wie Naturkundemuseen und Lebendsammlungen von Tieren, Pflanzen und Mikroorganismen sowie die Millionen von Beobachtungsdatensätzen, die z.B. zur Kartierung der Vegetation oder von Vogelzügen aufgenommen wurden.

GBIF-International und GBIF-Deutschland

GBIF ist eine internationale Einrichtung und besteht derzeit aus 47 Staaten und 30 internationalen Organisationen, die sich mit der Unterzeichnung einer Absichtserklärung dazu verpflichtet haben, „einen oder mehrere Knoten aufzubauen, die für GBIF zugänglich sind und Biodiversitätsdaten bereitstellen“ (GBIF 2000). Das Sekretariat in Kopenhagen koordiniert die einzelnen nationalen Knoten und bietet zentrale Dienste an.

Deutschland als Gründungsmitglied von GBIF hat mit Förderung des Bundesministeriums für Bildung und Forschung (BMBF) sieben Knoten eingerichtet, die sich thematisch an größeren Organismengruppen orientieren: Entomologie (Evertebraten I), Mollusken, Cheliceraten und Myriapoden (Evertebraten II), übrige marine Wirbellose (Evertebraten III), Prokaryonten, Botanik, Wirbeltiere und Mykologie. Insgesamt sind in Deutschland 66 Projekte aus 48 Forschungseinrichtungen am Aufbau der Knoten beteiligt. Ihre Aufgabe besteht sowohl darin, die Erfassung der in Deutschland vorhandenen Informationen zur Artenvielfalt zu koordinieren und diese in GBIF-International einzubringen als auch darin, die Nutzung von GBIF durch die deutsche Forschung zu fördern (Hahn u. a. 2004a).

Datentechnische Grundsätze von GBIF

Einer der wichtigsten Grundsätze ist die – aus Sicht von GBIF-International – dezentrale Datenhaltung: Der Bereitsteller der Information (Datenprovider) behält die vollständige Kontrolle über die von ihm gelieferten Daten. Es gibt keine zentrale Datenbank, in der die bereitgestellten Daten abgelegt werden müssen; stattdessen verbleiben sie in dem bestehenden System und werden dort bei Bedarf von GBIF abgefordert. Die Anbindung an GBIF hat somit weder Auswirkungen auf die gewohnte Arbeit mit der eigenen Sammlungsdatenbank, noch hat sie zur Folge, dass der Sammlungsinhaber seine Datenbestände an GBIF abgeben muss.

Die Entscheidung, welche Teile der Daten über GBIF zugänglich sind, liegt allein beim Provider. Der Anschluss an GBIF bedeutet nicht zwangsläufig, dass alle in der Sammlung vorhandenen Daten veröffentlicht werden müssen; vielmehr ist es möglich, gezielt Informationen auszublenden. So können ganze Datensätze gesperrt werden (um z.B. Beobachtungsdaten über gefährdete Taxa nicht zu veröffentlichen) oder auch nur Teilmeldungen eines Datensatzes (beispielsweise die exakten geografischen Koordinaten des Fundortes eines gefährdeten Taxons).

Der Zugriff auf die über GBIF veröffentlichten Daten erfolgt nicht zwangsläufig über ein zentrales Suchportal, das die Hoheit über die GBIF-Daten besitzt, sondern über potentiell zahlreiche Suchportale, die sich deutlich voneinander unterscheiden können. Das betrifft sowohl die Zielgruppe der Portale (Laien/Experten) als auch die umfassten Organismengruppen (z.B. ausschließlich Spinnentiere, Wirtspflanzen oder Vorratsschädlinge). Es steht jedem offen, das Netzwerk zu nutzen um ein neues Portal zu eröffnen, das z.B. eine bisher unterrepräsentierte Organismengruppe abdeckt oder das sich durch Layout, Bedienung und Funktionalität (z.B. durch Verknüpfung mit eigenen Daten) von den bestehenden Portalen unterscheidet.

Die Providersoftware („Wrapper“)

Der Grundsatz der dezentralen Datenhaltung im GBIF-Netzwerk wirft ein Problem auf: Die einzelnen Sammlungsdatenbanken unterscheiden sich zum Teil erheblich voneinander.

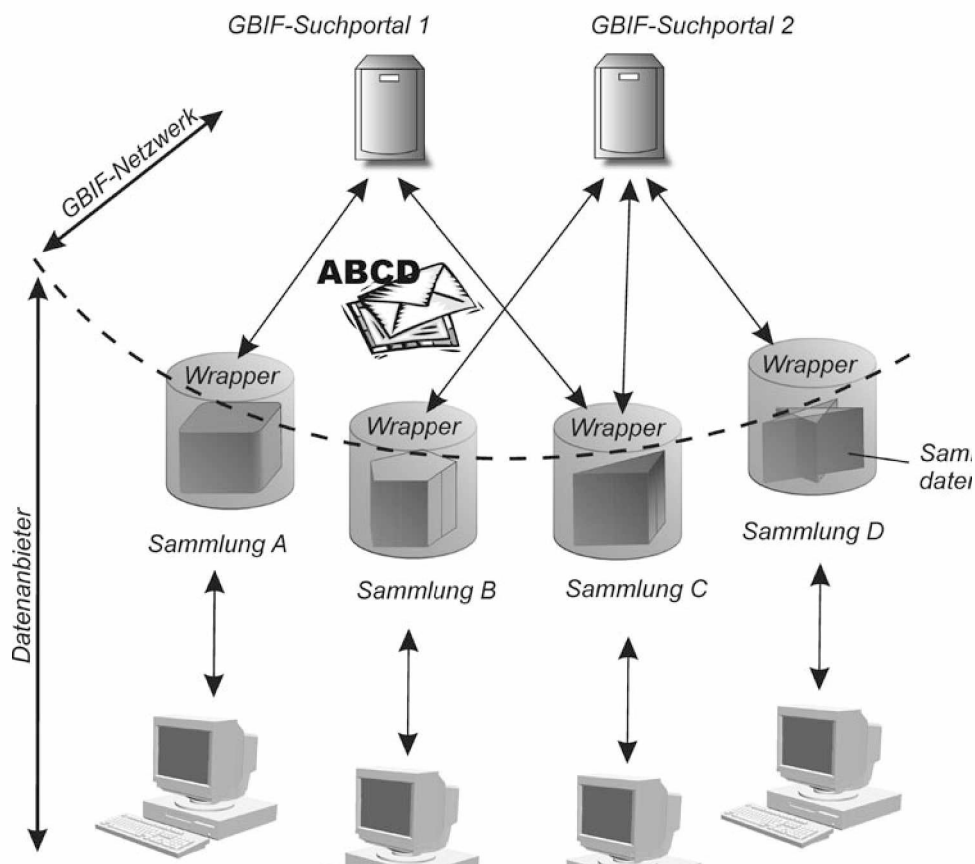


Fig. 1: Die Wrapper-Software umgibt die bestehenden Sammlungsdatenbanken mit einer zusätzlichen Abstraktionsschicht und bietet so eine definierte Schnittstelle zwischen den existierenden Datenbanksystemen und den GBIF-Suchportalen.

Dies betrifft sowohl die verwendeten Betriebssysteme (Windows, Linux, Unix, Solaris) und Datenbankmanagementsysteme (Access, MS SQL Server, Postgres, mySQL, Oracle, FileMaker, Sybase, FoxPro) als auch die Daten selbst – im Hinblick auf Umfang und Qualität wie auch hinsichtlich der Datenbankstruktur (Grad der Normalisierung). Trotz dieser Unterschiede muss es möglich sein, die Daten in einer einheitlichen Weise über das Internet abzufragen und die Ergebnisse in einem gemeinsamen Format geliefert zu bekommen.

Im Rahmen des BioCASE-Projektes (Biological Collection Access Service for Europe, Berendsohn u. a., 2000) wurden dazu Technologien und Standards sowie die entsprechende Software entwickelt (Güntsch u.a., 2005). Die Providersoftware („Wrapper“) umgibt das bestehende Datenbanksystem mit einer zusätzlichen Abstraktionsschicht und bietet so eine definierte Schnittstelle zu den Sammlungsdaten (Fig. 1). Dieser Wrapper übernimmt zwei Aufgaben: Zum einen nimmt er die Anfragen der Suchportale entgegen und übersetzt sie in eine für das Datenbanksystem verständliche Sprache; zum anderen verpackt er die gelieferten Ergebnisse in ein für die Suchportale verständliches Datenschema (Access to Biological Collection Data, ABCD).

Die Providersoftware kann mit allen SQL-fähigen Datenbanksystemen umgehen und wurde in der Skriptsprache Python entwickelt, die für alle gängigen Betriebssysteme verfügbar ist. Bisher konnte der Wrapper auf allen existierenden Betriebssystem/Datenbankmanagementsystem-Kombinationen erfolgreich eingesetzt werden.

Datenflüsse im GBIF-Netzwerk

Figur 2 zeigt die Datenflüsse bei einer Suche im GBIF-Netzwerk. Vereinfachend ist nur ein Datenlieferant auf der linken Seite dargestellt, bestehend aus Sammlungsdatenbank und Wrapper-Software. Im realen Netzwerk existieren natürlich zahlreiche Daten-Provider. Ebenso ist nur eines der existierenden Suchportale gezeigt, in der rechten, unteren Seite. Die Verbindung zwischen Datenlieferant, Suchportal und Nutzer erfolgt über das Internet.

Der Nutzer, der Biodiversitätsdaten im GBIF-Netzwerk suchen möchte, verwendet seinen Internetbrowser, um auf das Suchportal zuzugreifen und über eine Suchmaske seine Suchanfrage einzugeben (z.B. nach einem Artnamen). Die ebenfalls im BioCASE-Projekt entwickelt Software *UnitLoader* verschickt nun Suchanfragen an alle infrage kommenden Datenlieferanten; welche das sind, wird mit Hilfe eines Verzeichnisses aller verfügbaren Datensätze ermittelt (Index). Die Wrapper-Software nimmt die Anfrage entgegen und leitet sie an die Sammlungsdatenbank weiter, die ihrerseits die der Suche entsprechenden Ergebnisse zusammenstellt. Der Wrapper verpackt die Ergebnisse im ABCD-Datenschema und sendet sie an das Suchportal zurück. Dieses wartet, bis die Antworten aller verfügbaren Daten-Provider eingetroffen sind und stellt anschließend die Ergebnisse für den Nutzer dar. Dieser Mechanismus kann natürlich auch – an Stelle des Benutzers – von einer Software genutzt werden, z.B. für statistische Analysen oder Modellierung mittels geographischer Informationssysteme.

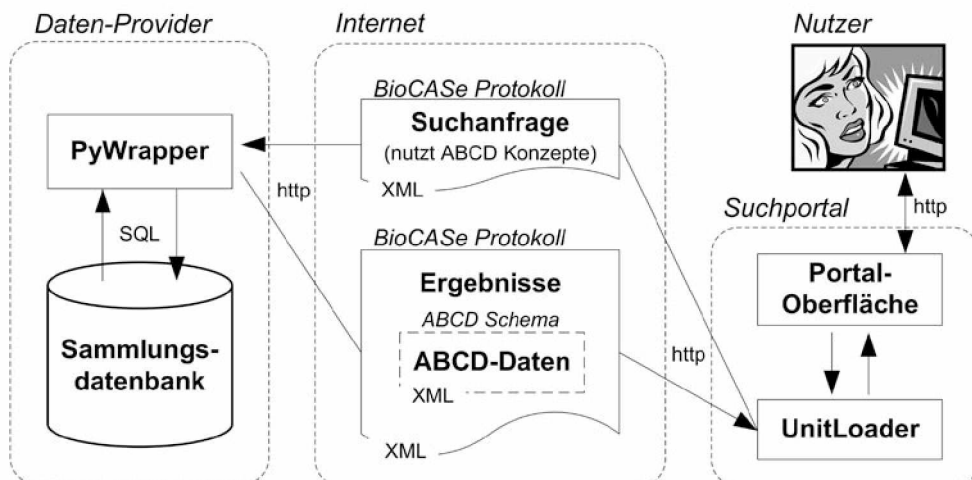


Fig. 2: Datenflüsse über das Internet im GBIF-Netzwerk zwischen Nutzer, Suchportal und Datenlieferant.

BioCASE-Protokoll und ABCD-Datenschema

Sowohl die vom Portal versendete Suchanfrage als auch die zurückgelieferten Ergebnisse werden nach dem BioCASE-Protokoll übertragen (Döring & Güntsch 2005). Das Protokoll regelt die Formalitäten der Suche im Netzwerk, also beispielsweise die Art der Suche, die Maximalzahl der gewünschten Treffer oder Timeout-Werte, mit deren Hilfe die zeitliche Dauer des Suchvorgangs begrenzt werden kann. Ein weiteres im GBIF-Netzwerk verwendetes Protokoll ist DiGIR (Distributed Generic Information Retrieval). Für zukünftige Entwicklungen wurde bereits festgelegt, beide Protokolle zu einem gemeinsamen Standard zu verschmelzen.

Die Biodiversitätsdaten selber – also Beobachtungs- oder Beleginformationen – werden konform zum ABCD-Datenschema abgelegt (Berendsohn 2005), das Syntax und Semantik der übertragenen Informationen festlegt. ABCD ist XML-basiert (eXtensible Markup Language) und ermöglicht die Strukturierung unterschiedlichster biologischer Sammlungsdaten (Beobachtungen/Belege, lebend/konserviert, botanisch/zoologisch/mikrobiologisch, marin/terrestrisch). In der neuesten Version 2.0 enthält ABCD 1.055 Datenfelder zur Speicherung der Informationen und zielt auf die vollständige Beschreibung der Sammlungsdaten, sowohl der eigentlichen Beobachtungs-/Belegdatensätze („units“) als auch der Metadaten, also der Angaben zu Sammlung, Sammlungsbetreuer, der bereitstellenden Institution und Urheberrechten. Selbstverständlich müssen für eine spezielle Sammlung nicht alle ABCD-Datenfelder geliefert werden; bis auf wenige Ausnahmen ist ihre Angabe optional. ABCD ist hierarchisch gegliedert und kann verschieden stark atomisierte, d.h. sich im Grad ihrer Strukturierung unterscheidende Daten aufnehmen.

Mapping

Ein ABCD-Dokument besitzt eine hierarchisch gegliederte Baumstruktur, in deren einzelnen Elementen verschiedene Informationen abgelegt werden können. Die Sammlungsdatenbank speichert die Informationen in einzelnen, flachen Tabellen und Spalten (relationales Datenschema). Diese beiden „Welten“ müssen aufeinander abgebildet werden (Fig. 3). Dieser Prozess wird als Mapping bezeichnet und stellt die Hauptaufgabe bei der Anbindung einer Sammlung an GBIF dar. Konkret müssen die einzelnen Datenbankspalten der Sammlungsdatenbank den entsprechenden ABCD-Elementen zugeordnet werden. Dies ermöglicht es dem Wrapper später, eine in ABCD-Elementen formulierte Suchanfrage in eine für die Datenbank verständliche Form zu übersetzen und, auf umgekehrtem Wege, die Suchergebnisse der Datenbank ins ABCD-Datenschema zu konvertieren.

Der Mappingvorgang erfolgt in einem komfortablen Konfigurationstool, das Teil der Providersoftware ist. Dieses ist browserbasiert und kann somit auch aus der Ferne erfolgen (Remote-Konfiguration). Eventuelle Filterregeln, die Teile der Daten nach bestimmten Kriterien ausblenden sollen, werden ebenfalls während des Mappings festgelegt.

Direkte Anbindung einer Sammlungsdatenbank an GBIF

Zur unmittelbaren Anbindung einer bestehenden Sammlungsdatenbank an das GBIF-Netzwerk müssen zwei Bedingungen erfüllt sein: Zum einen ist ein dauerhaft mit dem Internet verbundener Webserver nötig, was bei den meisten größeren Einrichtungen der Fall ist. Die zweite Voraussetzung betrifft das Datenbankmanagementsystem, in dem sich die Sammlungsdaten befinden. Es muss die Abfragesprache SQL (Structured Query Language) beherrschen; dies ist bei allen gängigen Systemen der Fall.

Nachdem diese Voraussetzungen sichergestellt sind, erfolgt die Anbindung in vier Schritten:

1. Installieren der Providersoftware und der Python-Skriptsprache.
2. Vorbereiten der Sammlungsdatenbank für die Anbindung: Festlegung der zu veröffentlichenden Informationen, eventuell Export dieser Daten in eine zweite Instanz der Datenbank.
3. Konfiguration der Providersoftware; Durchführen des Mappings.
4. Testen der neuen Datenquelle und Registrierung im GBIF-Netzwerk.

Der erste und letzte Schritt sind rein technischer Natur und können selbständig von den jeweiligen EDV-Mitarbeitern durchgeführt werden. Schritt zwei und drei hingegen erfordern die Mitarbeit des Sammlungsbetreuers.

GBIF-Deutschland ist den Sammlungsbetreuern bei der Anbindung ihrer Datenbank an GBIF behilflich. Am Botanischen Museum in Berlin existiert ein Helpdesk, der bei Unklarheiten erklärend zur Seite steht oder bei konkreten Problemen der Umsetzung hilft. Dies ist telefonisch oder per E-Mail, mittels Remote-Konfiguration (besonders beim Mappingvorgang) oder in schwierigen Fällen auch direkt vor Ort möglich. Diese Unterstützung ist bis Dezember 2006 über Projektfinanzierungen gesichert.

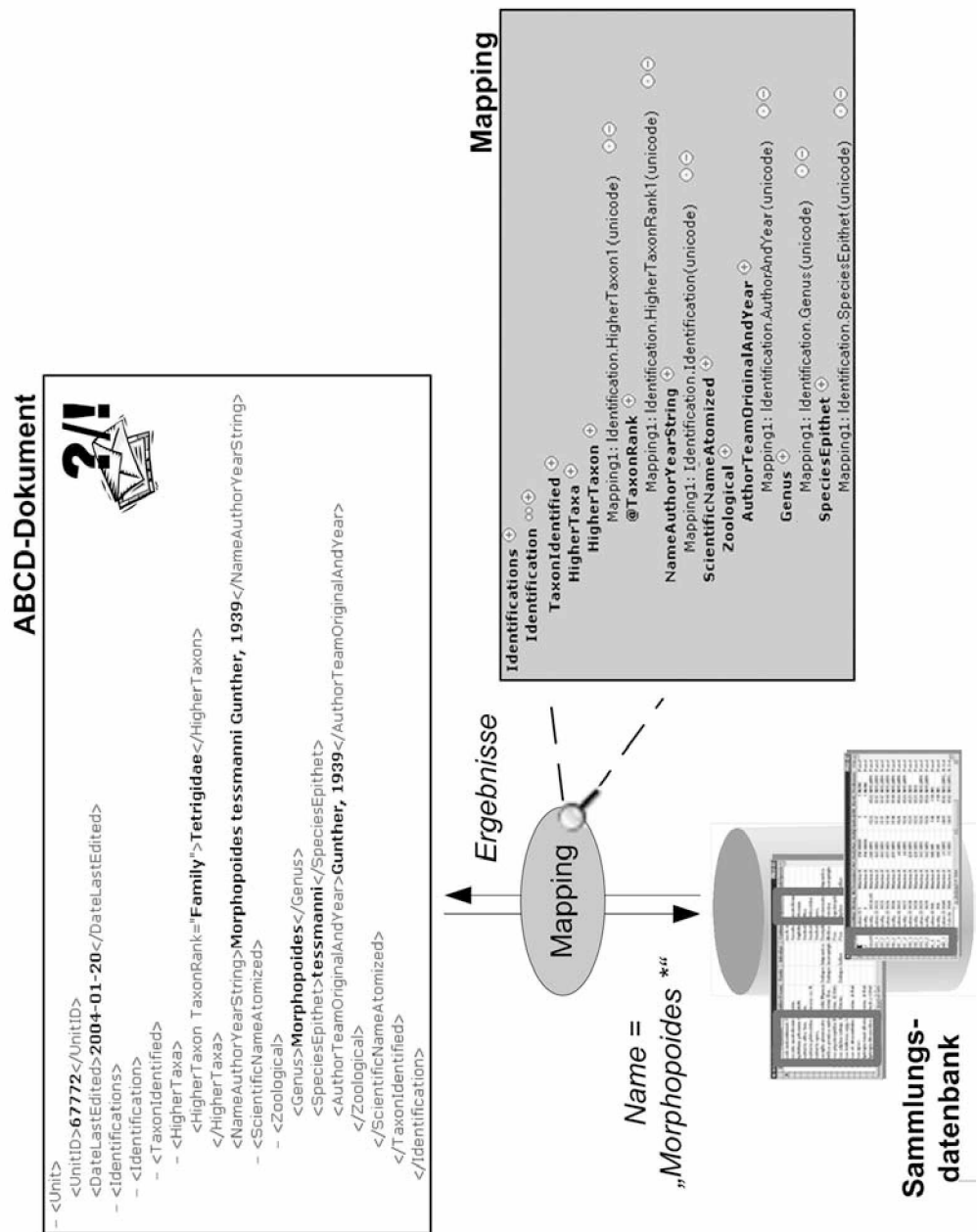


Fig. 3: Die Zuordnung der Daten aus dem relationalen Datenschema der Sammlungsdatenbank zu den ABCD-Elementen wird im Mapping festgelegt und ist in einer komfortablen Oberfläche mit dem Internet-Browser möglich.

Alternative Anbindungsmöglichkeiten an GBIF

Neben dem soeben beschriebenen Weg, ein unmittelbarer Datenlieferant zu werden, existieren alternative Möglichkeiten zur Einbindung von Datenbeständen in GBIF. Diese können vor allem für kleinere Sammlungen in Betracht kommen, die nicht über die notwendige technische Infrastruktur verfügen oder aber deren Datenmengen so gering sind, dass sich der mit einer direkten Anbindung verbundene Aufwand nicht lohnt. Für einige wenige Typusdatensätze mag die Einrichtung eines eigenen GBIF-Providers unangemessen sein; gleichwohl ist es wichtig, die Daten verfügbar zu machen.

Für diese Fälle gibt es zwei alternative Lösungsmöglichkeiten:

- Die Sammlungsdatenbank kann einem bereits bestehenden Provider zur Verfügung gestellt werden, der schon über die notwendige Infrastruktur und Software verfügt und die Sammlung stellvertretend für den Sammlungsinhaber anbindet. Die Sammlung bleibt weiterhin als eigenständige Datenbank erhalten, und als Quelle der Informationen in GBIF wird weiterhin der Sammlungsinhaber genannt werden. Die anbindende Institution ist lediglich der Anbieter im technischen Sinne.
- Die Daten können – zusammen mit den Daten aus anderen Quellen – in einer gemeinsamen zentralen Datenbank gehalten werden, die ihrerseits an GBIF angeschlossen ist. Die Daten werden entweder in das gemeinsame System importiert oder über entsprechende Clients direkt in diesem System bearbeitet.

Erstere Variante wird derzeit im Botanik-Knoten von einigen Herbarien genutzt. Beispiele für die in Variante 2 erwähnte gemeinsame Datenbank sind das von der Universität Ulm und der Ruhr-Universität Bochum entwickelte Datenbanksystem für Systematik und Taxonomie (SysTax, Hoppe u.a. 2005) sowie das Sammlungsverwaltungssystem SeSam des Forschungsinstituts und Naturkundemuseums Senckenberg in Frankfurt (Türkay u.a. 2005).

Ergebnis

Derzeit (Stand: Juni 2005) haben international bereits 126 Datenanbieter Informationen aus 514 Sammlungen und Beobachtungsdatenbanken an GBIF angeschlossen und stellen so über die Suchportale von GBIF mehr als 74 Millionen Datensätze zur Verfügung (Fig. 4). Jedoch sind damit bei weitem noch nicht alle vorhandenen elektronischen Daten online verfügbar; ein Großteil wartet noch auf seine Erschließung. Jeder Betreuer einer Sammlung mit Beobachtungs- oder Belegdaten hat die Möglichkeit, durch die Anbindung seiner Daten an GBIF die Erreichbarkeit seiner Primärdaten deutlich zu erhöhen und darüber hinaus die Datenbasis zu vergrößern, die der Wissenschaft und den Entscheidungsträgern in Politik und Umweltschutz zur Verfügung steht.

Bereits jetzt existieren mehrere Suchportale, mit deren Hilfe die Daten im GBIF-Netzwerk durchsucht werden können:

- Die zentrale GBIF-Webseite unter www.gbif.org enthält neben einer namens- und geografiebasierten Suche eine umfangreiche taxonomische Datenbank, in der geblättert werden kann.

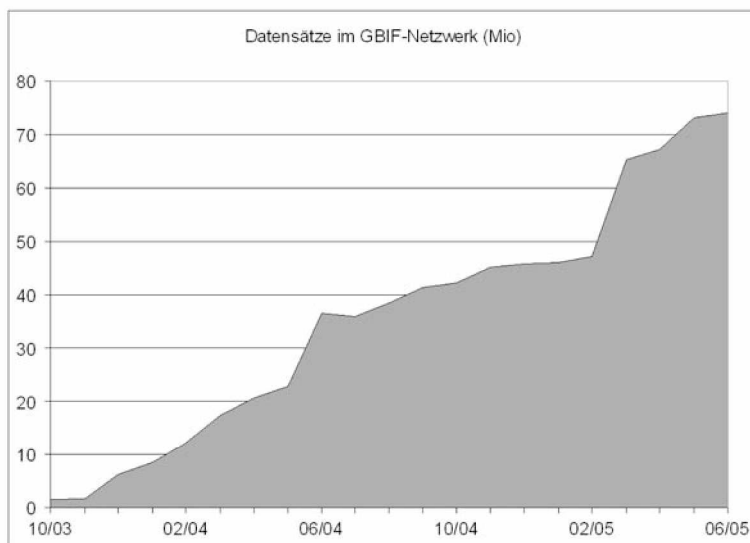


Fig. 4: Anzahl der im GBIF-Netzwerk verfügbaren Datensätze (Belege und Beobachtungen) in Millionen.

Botanische Sammlungen in Deutschland und Virtuelles Herbarium
- Prototyp -

Bitte beachten Sie:
Dies ist ein Prototyp, um den verteilten Zugriff auf ABCD-Datenquellen unter Benutzung von BioCASE-Protokoll und -Software zu demonstrieren. Das endgültige Portal befindet sich in Entwicklung. Klicken Sie hier für weitere Informationen zu diesem Prototyp (in Englisch).

☆ **Suchanfrage an ABCD-Provider:**
Geben Sie Ihre Anfrage ein - benutzen Sie bitte * als Platzhalter. Im wissenschaftlichen Namen wird am Ende vom System automatisch ein Platzhalter ergänzt (Beispiel: als Suchkriterium im Wissenschaftlichen Namen liefert "Abi" alle Namen, die mit Abi beginnen, "Abies alba" alle Unterarten, Varietäten etc. von Abies alba, "*aaa*" findet alle Namen, die ein dreifaches a enthalten, "Ab*ba" findet unter anderem Abies alba etc.). Wenn Sie die vollständigen Resultate in XML-Format zugeschickt bekommen möchten, geben Sie bitte eine gültige Email-Adresse ein.

Geben Sie einen Wissenschaftlichen Namen ein (z.B. Hierac):	<input type="text" value="Abies alba"/>
Geben Sie ein Höheres Taxon ein (z.B. P*):	<input type="text" value="*"/>
Geben Sie einen Ländernamen ein (englischsprachig, z.B. Iceland):	<input type="text" value="*"/>
Geben Sie zur Zusendung der vollständigen Ergebnisse Ihre Email-Adresse an:	<input type="text" value="*"/>

Maximale Anzahl von Ergebnissen pro Datenbank:
Nicht mehr als Sek. auf die Antwort jeder Datenbank warten.

Powered by:

© 2005 Freie Universität Berlin, Botanischer Garten und Botanisches Museum Berlin-Dahlem
Seitenverantwortliche/Page editors: J. Holetschek, A. Hahn, J. de la Torre, A. Güntsch
BGBM Impressum/Imprint; Haftungsausschluss/Disclaimer

Fig. 5: Das Suchportal des Botanik-Knotens von GBIF-Deutschland, einer der mehreren im Internet verfügbaren Zugangspunkte zu den Daten des GBIF-Netzwerks.

- Das Suchportal des BioCASE-Projektes unter search.biocase.org ermöglicht die Suche nach einer Vielzahl von Kriterien.
- Der Botanik-Knoten von GBIF-Deutschland (Hahn & al. 2004b) besitzt ein einfaches Suchportal (Fig. 5), mit dem die deutschen botanischen Biodiversitätsdaten durchsucht werden können (www.gbif.de/botanik).

Im Rahmen der Knotenstruktur existieren des weiteren zwei Portale, die Zugang zu Knotendaten auf Basis ihrer eigenen Netzwerkarchitektur und Datenstruktur ermöglichen:

- Mit dem SysTax Zoologie-Portal ist es möglich, jeweils auf alle Datenlieferanten der Knoten Evertibraten I und II sowie Wirbeltiere unter einer gemeinsamen Webseite zuzugreifen (für die Entomologie: www.biologie.uni-ulm.de/cgi-bin/portal/portal.pl?lang=d&pr=gbif-el).
- Das SeSam-Portal unter www.senckenberg.de/gbif_welcome dient zur Suche im Knoten Evertibraten III.

Ausblick

Eine große Herausforderung für GBIF ist die Menge der Daten, die bisher noch nicht angeschlossen wurden. Bisher wurde nur ein geringer Teil der weltweiten Belege digitalisiert, und die Namenskataloge sind keineswegs vollständig. Zwar schwinden die anfänglichen Vorbehalte der Sammlungsbetreuer und Mitarbeiter in dem Maße, in dem der Nutzen der von anderen verfügbar gemachten Daten deutlich wird, jedoch steht bis zur Erreichung des Ziels noch ein weiter Weg bevor.

Mit wachsenden Datenmengen wird auch das Problem der Datenqualität an Bedeutung gewinnen. In der Anfangsphase war es vorrangiges Ziel von GBIF, neue Datenquellen zu erschließen, um eine möglichst vollständige Abdeckung aller Organismengruppen zu erreichen. Die Qualität dieser Daten variiert stark, von aus Monografien entnommenen Checklisten bis zu unvollständigen und unverifizierten Sammlungen (Berendsohn 2004). Das GBIF-Arbeitsprogramm DIGIT (Digitisation of Natural History Collections) beschäftigt sich mit der Entwicklung effizienter Digitalisierungsmethoden und wirksamer Wege zur Qualitätssicherung (GBIF 2005a).

Das abzusehende Datenwachstum stellt GBIF vor enorme technische Probleme. Allein in der EU existieren über 5,000 potentielle Datenanbieter. Der mit zunehmender Informationsabdeckung einhergehende Anstieg der Suchanfragen dürfte vor allem für kleinere Datenanbieter nicht zu bewältigen sein. Die Notwendigkeit der zentralen Zwischenspeicherung der Daten (Caching) wird somit immer größer. Dafür müssen Methoden entwickelt werden, um die Aktualisierung der Daten zu gewährleisten; ebenso muss es dem Provider möglich sein, seine Daten aus dem Zwischenspeicher zu löschen oder deren vollständige Aktualisierung zu erzwingen. GBIF befasst sich im Rahmen des Arbeitsprogramms DADI (Data Access and Data Interoperability) mit diesen Fragen (GBIF 2005b).

Weiterentwicklungsbedarf besteht auch bei den Schnittstellen zum Nutzer (Suchportalen). Zum einen sollten sie in Zukunft die Biodiversitätsdaten mit Informationen aus anderen Bereichen verknüpfen können (beispielsweise Thesauri), andererseits muss der Benutzer von der wachsenden Komplexität der Biodiversitätsdaten abgeschirmt werden.

Weiterhin müssen die Suchportale die zu erwartenden Nutzerzahlen bedienen können; die Einrichtung paralleler Websites (Mirroring) und deren Datenabgleich untereinander wird notwendig. Mehrere, aus verschiedenen Quellen finanzierte Projekte befassen sich mit diesen Problemen.

Weitere Entwicklungen betreffen die im GBIF-Netzwerk verwendeten Protokolle und Datenschemata. BioCASE- und DiGIR-Protokoll werden zu TAPIR verschmelzen (TDWG Access Protocol for Information Retrieval, TDWG = Taxonomic Databases Working Group, Döring & de Giovanni 2004). Das ABCD-Datenschema, welches im September der TDWG zur Verabschiedung als Standard vorgelegt werden soll, wurde in seiner jetzigen Version (2.0) bereits um spezifische Elemente – vor allem für Kulturpflanzensammlungen und mikrobiologische Kulturen – erweitert. Dies hat Auswirkungen sowohl auf die Suchportale als auch auf die bei den Providern installierte Wrapper-Software. Eine Weiterentwicklung hinsichtlich noch breiterer Anwendbarkeit ist vorgesehen, vermutlich wird ABCD auf lange Sicht zusammen mit anderen TDWG-Standards (beschreibende Daten, Taxon-Namen und Konzepte, etc.) in einzelne Objekte modularisiert werden.

Wichtigster Erfolgsfaktor für GBIF bleibt jedoch die verfügbare Datenbasis. Die weitere Mobilisierung von Biodiversitätsdaten ist unerlässlich und benötigt die Unterstützung der Besitzer primärer Biodiversitätsinformationen.

Referenzen

- BERENDSOHN, W. G. 2004: The German contribution to the Global Biodiversity Information Facility: GBIF-D and BIOLOG Biodiversity Informatics. Pp. 176-185 in Beck & al. (ed.): Sustainable use and conservation of biological diversity – A challenge for society. - Proceedings of the Symposium Berlin 1-4 December 2003. Bonn.
- BERENDSOHN, W. G. 2005: ABCD 2.0, the proposed TDWG standard for Access to Biological Collection Data. - Berlin. [<http://www.bgbm.org/TDWG/CODATA/Schema/>]
- BERENDSOHN, W. G.; COSTELLO, M. J.; EMBLOW, C.; GÜNTSCH, A.; HAHN, A.; KOENEMANN, J.; THOMAS, C.; THOMSON, N. & WHITE, R. 2000: Concepts for a European Portal to Biological Collections. - In: BERENDSOHN, W.G. (ed.): Resource Identification for a Biological Collection Information Service in Europe, Botanic Garden and Botanical Museum Berlin Dahlem.
- DÖRING, M. & GÜNTSCH, A. 2005: The BioCASE unit-level access protocol. - Berlin. [<http://www.biocase.org/dev/protocol/index.shtml>]
- DÖRING, M. & DE GIOVANNI, R. 2004: A unified protocol for search and retrieval of distributed data. - GBIF Data Access and Database Interoperability Workprogramme. [<http://www.cria.org.br/protocols/newprotocol.pdf>]
- GBIF 2000: Memorandum of Understanding for the Global Biodiversity Information Facility. - GBIF Secretariat Copenhagen. [http://www.gbif.org/GBIF_org/GBIF_Documents/mou_html]
- GBIF 2005a: The Digitisation of Natural History Collections Workprogramme (DIGIT). - GBIF Secretariat Copenhagen. [<http://www.gbif.org/prog/digit>].
- GBIF 2005b: The Data Access and Data Interoperability (DADI). - GBIF Secretariat Copenhagen. [<http://www.gbif.org/prog/dadi>].
- GÜNTSCH, A.; DÖRING, M. & BERENDSOHN, W. G. 2005 (in Druck): Mobilisierung von primären Biodiversitätsdaten: Das BioCASE Protokoll und seine Anwendung in internationalen Netzwerken. - Workshop Umweltdatenbanken der GI-Fachgruppe 4.6.1 Informatik im Umweltschutz der Gesellschaft für Informatik e.V., Hannover, 6.-7. Juni 2005. [<http://www.umwelt.schleswig-holstein.de/?AKUmweltdatenbanken>].

- HAHN, A.; KIRCHHOFF, A. & BERENDSOHN, W. G. 2004a: The Global Biodiversity Information Facility (GBIF) and the German response. - *Scripta Botanica Belgica*, vol. 29, pp. 71-77, Meise.
- HAHN, A.; KIRCHHOFF, A. & BERENDSOHN, W. G. 2004b: Virtual Herbarium, portal and organisation of the Botanical Node within GBIF-D. Pp. 48-49. - In: BERENDSOHN, W. G. & OEHLSCHLAEGER, S. (ed.): GBIF-Deutschland - German participation in the Global Biodiversity Information Facility. - Status Report 2004. Bonn & Berlin.
- HOPPE, J. R.; BOOS, E.; LUDWIG, T.; WIEDEMANN, M. & STÜTZEL, T. 2005: A Database System for Systematics and Taxonomy. - Ulm, Bochum. [<http://www.biologie.uni-ulm.de/systax>]
- OEHLSCHLAEGER, S. & BERENDSOHN, W. G. 2005: Die Global Biodiversity Information Facility. - GBIF-Deutschland, Berlin.
- TÜRKAY, M.; ALLSPACH, A. & MENNER, L. 2005: Sammlungsverwaltungssystem SeSam. - Frankfurt. [<http://www.senckenberg.de/sesam>]

Anschrift der Verfasser:

Dipl. Wirt.-Inf. JÖRG HOLETSCHEK und Prof. Dr. WALTER G. BERENDSOHN
Botanischer Garten & Botanisches Museum Berlin-Dahlem,
Abteilung Biodiversitätsinformatik und Labors, Königin-Luise-Straße 6-8, D- 14191 Berlin
e-mail: j.holetschek@bgbm.org und w.berendsohn@bgbm.org

ZOBODAT - www.zobodat.at

Zoologisch-Botanische Datenbank/Zoological-Botanical Database

Digitale Literatur/Digital Literature

Zeitschrift/Journal: [Beiträge zur Entomologie = Contributions to Entomology](#)

Jahr/Year: 2005

Band/Volume: [55](#)

Autor(en)/Author(s): Berendsohn Walter G., Holetschek Jörg

Artikel/Article: [Freier Zugang zu den Informationen der Artenvielfalt - Wie werde ich Teil der Global Biodiversity Information Facility \(GBIF\)? 433-444](#)