

# **Abschlussbericht** der **IT-Fachgruppe (GBIF-D)**

**Januar 2004 bis Dezember 2007**

Workshops zur Informationstechnik  
im Bereich Biodiversitätsinformatik

---

Dokumentation und Analyse von IT-Entwicklungen





## **Abschlussbericht der IT-Fachgruppe (GBIF-Deutschland)**

---

Projekttitle:	<b>Koordination der IT-Fachgruppe für GBIF-D</b>
Förderkennzeichen:	<b>01LI0301</b>
Zuwendungsempfänger:	<b>Prof. Dr. G. Rambold, Universität Bayreuth</b>
Vorhabensbezeichnung:	<b>Koordination einer IT- Fachgruppe, die den Informationsaustausch und die Zusammenarbeit zwischen Software- entwicklern gewährleistet und fördert.</b>
Laufzeit des Vorhabens:	<b>1.1.2004 bis 31.10.2007</b>
Berichtszeitraum:	<b>1.1.2004 bis 31.12.2007</b>

---

Koordination: Prof. Dr. G. Rambold  
Koordinationsassistentz: Dr. A. Kehl  
Website: [http://132.180.63.25/GBIF-ITFG/ITFG\\_welcome.html](http://132.180.63.25/GBIF-ITFG/ITFG_welcome.html)

## **Kontakt**

Dr. Alexandra Kehl  
Universität Bayreuth  
Universitätsstraße 30  
D-95440 Bayreuth  
Tel.: +49 921 552455  
Fax: +49 921 552567  
E-Mail: [alexandra.kehl@uni-bayreuth.de](mailto:alexandra.kehl@uni-bayreuth.de)

Prof. Dr. Gerhard Rambold  
Universität Bayreuth  
Universitätsstraße 30  
D-95440 Bayreuth  
Tel.: +49 921 552453  
Fax: +49 921 552567  
E-Mail: [gerhard.rambold@uni-bayreuth.de](mailto:gerhard.rambold@uni-bayreuth.de)

## INHALTSVERZEICHNIS

1. Fazit.....	1
2. Kurze Darstellung des Vorhabens .....	2
3. Durchgeführte Workshops .....	3
3.1. Übersicht .....	3
3.2. Bilanz.....	5
4. Dokumentation der IT-Entwicklungen.....	6
4.1. Einleitung.....	6
4.2. Zusammenschau der Ergebnisse .....	7
4.3. Interpretation der Ergebnisse .....	15
5. Website der IT-Fachgruppe .....	16
6. Danksagung .....	17
7. Anlagen .....	17
Anlage 1: Protokolle zu den Workshops .....	17
Anlage 2: Gesamtliste der Workshopteilnehmer.....	101
Anlage 3: Gesamtliste der Workshop-Beiträge.....	103
Anlage 4: Merkmalslisten der HTML-Formulare .....	104
Anlage 5: Auflistung der dokumentierten IT-Entwicklungen .....	111
Anlage 6: Beschreibung der dokumentierten IT-Entwicklungen .....	114

Blau hinterlegte Abschnitte kennzeichnen wesentliche Ergebnisse bzw. Zusammenfassungen.

## 1. Zusammenfassung – Conclusions

An den Workshops der GBIF IT-Fachgruppe nahmen sowohl Wissenschaftler als auch Informatiker und Softwareentwickler teil. Die hohe Teilnehmerzahl (insgesamt 69) und die Resonanz zu den Workshops zeigen, dass die Möglichkeit an solchen Treffen teilzunehmen, sehr positiv aufgenommen wurde. Die Einrichtung der IT-Fachgruppe in der bestehenden Form wurde mehrheitlich als gut und sinnvoll erachtet.

Aus Sicht der IT-Fachgruppe wird vor allem für Projekte aus der Forschung die Nutzung bereits bestehender IT-Strukturen und Entwicklungen als günstig angesehen. Um die Verknüpfung der Datenflüsse zu gewährleisten, sollten auch eigene Projekte zur Koordination und Entwicklung von IT-Strukturen im Rahmen von Biodiversitäts-Großprojekten beantragt werden.

Im Bereich der Biodiversitätsforschung wird die Etablierung von informationstechnischen Datennetzwerken gefordert. Es können dafür u. a. Systeme und Schnittstellen genutzt werden, die GBIF bereitstellt.

Aus Sicht der Mitglieder der IT-Fachgruppe kann die Phase 1 von GBIF-Deutschland im Bezug auf die Etablierung von Strukturen und IT-Techniken als Erfolg betrachtet werden. Von zahlreichen Teilnehmern wurde der Wunsch nach einer Weiterführung des Projekts geäußert.

The meetings of the GBIF IT-expert group were attended by natural scientists as well as by computer scientists and software developers. The high number of participants (all together 69) and their positive reactions showed that the possibility to take part in such meetings were highly appreciated. The majority of participants regarded the establishment of the IT-section in its present form as good and reasonable.

The use of already existing IT structures and developments i.e. for research projects is considered as advantageous. In order to provide for the linkage of data flows, one should also apply for own (special) projects of coordination and development of IT-structures within the framework of major biodiversity projects.

In the field of biodiversity research the establishment of IT-supported data networks is postulated. There would be the possibility to use i.a. systems and interfaces made available by GBIF.

According to the members of the IT-expert group, phase 1 of GBIF-Germany was a success with regard to the establishment of structures and IT-techniques. A number of participants expressed their desire for a prosecution (installment) of the project.

## 2. Kurze Darstellung des Vorhabens

Die IT-Fachgruppe war Teil der Koordinationsstruktur von GBIF-Deutschland neben der Gruppe der Knotenkoordinatoren und der IT-Kommission. Sie wurde als Informationsschnittstelle auf datentechnischer Ebene für Softwareentwickler konzipiert. Die Aufgabe der IT-Fachgruppe bestand sowohl in der Zusammenführung aller verfügbaren technischen Informationen über existierende Software- bzw. Datenbankentwicklungen als auch in der Koordination des Informationsaustausches der im Bereich der Biodiversitätsinformatik aktiven Softwareentwickler im Rahmen von Workshops. Die besonders enge Zusammenarbeit zwischen IT-Fachgruppe und IT-Kommission sollte dazu führen, dass Empfehlungen an die IT-Kommission erarbeitet werden, sowie dort aufgeworfene Fragen in der IT-Fachgruppe bearbeitet werden sollten. Auf den Workshops sollte den eingeladenen IT-Fachleuten bzw. Softwareentwicklern Gelegenheit geboten werden, technische Details ihrer IT-Entwicklungen vorzustellen und die anderer Projekte kennen zu lernen. Dabei sollten sowohl überlappende Aktivitäten und Redundanzen aufgedeckt, als auch die Entwicklung gemeinsamer plattformunabhängig einsetzbarer Anwendungen oder Webschnittstellen abgestimmt werden.

### **Workshops**

Workshops sollten nach Bedarf mehrmals im Jahr organisiert und vorbereitet werden. Die Teilnehmer der Workshops stellen die Mitglieder der IT-Fachgruppe dar. Ergebnisse der Workshops werden neben den Teilnehmern auch den Mitgliedern der IT-Kommission (GBIF-Deutschland) zur Verfügung gestellt und dienen als Hintergrundinformation und Dokumentationsmaterial für Anträge. Protokolle der Workshops sollten für die Allgemeinheit zugänglich auf einer Website hinterlegt werden. Ein wesentliches Ziel der Workshops bestand in der Unterstützung des Informationsaustausches zwischen Softwareentwicklern und IT-Verantwortlichen verschiedener Projekte. Hierbei sollten vor allem technische Aspekte im Vordergrund stehen, Evaluationen bzw. Stärken-Schwächen-Analyse durchgeführt sowie Lösungsansätze und Änderungsmöglichkeiten herausgearbeitet werden.

### **Dokumentation von IT-Entwicklungen**

Anhand laufend aktualisierter Bestandsaufnahmen sollte innerhalb von 3 Jahren eine detaillierte Dokumentation der IT-Strukturen und -Entwicklungen im Bereich GBIF-Deutschland und BIOLOG erstellt werden, ergänzt durch die Darstellung von eigenen IT-Entwicklungen, die außerhalb der BMBF-Förderung angesiedelten Projekten erstellt sind.

## Website

Die Informationen und Protokolle zu den Workshops sollten den Teilnehmern und der Öffentlichkeit über eine Website zugänglich gemacht werden. Zugleich konnte die Abfrage zu IT-Entwicklungen über Online-Formulare über die Website möglich gemacht werden.

## Mitglieder

Die IT-Fachgruppe sollte offen sein für alle Fachleute, die mit der Implementierung von Komponenten des deutschen GBIF-Systems oder anderen, zur Informationsbereitstellung im GBIF System genutzten Projekten direkt befasst sind.

## 3. Durchgeführte Workshops

### 3.1. Übersicht

Die Daten, Veranstaltungsorte und Themen der durchgeführten Workshops der IT-Fachgruppe sind der nachfolgenden Liste zu entnehmen. Die jeweiligen Protokolle sind in Anlage 1 zu finden.

Nr.	Datum	Thema	Ort
I	23.-24.4.2004	„Sammlungsdatenbanksysteme“	BSM, München
II	28.-29.10.2004	„Spezifikationen zum Datentransfer und zukünftige Datenflußszenarien.“	BGBM, Berlin
III	6.-7.1.2005	„Konzept für eine Rich Client Plattform zur Integration von <i>DiversityWorkbench</i> Komponenten.“	BSM, München
IV	12.-13.5.2005	„Integration und Anbindung von BIOTA-AFRICA Forschungsdaten in das GBIF-Netzwerk.“	BSM, München
V	22.-23.6.2005	„Erstellung eines Datenmodells für die in BIOTA erfassten Standorteigenschaften (Klimatologie, Bodenkunde, Fernerkundung).“	Universität Würzburg
VI	14.-15.7.2005	„Datenbanken zur Verwaltung ökologischer Daten in BIOTA und GBIF am Beispiel botanischer und zoologischer Beobachtungsdaten: Anforderungen, Gemeinsamkeiten, Perspektiven.“	FIS, Frankfurt a. M.
VII	29.-30.9.2005	„Für den Einsatz in BIOTA AFRICA geeignete Datenbanksysteme und Clients; Notwendigkeit BIOTA-spezifischer Anpassungen; Möglichkeiten bzw. Strategien für Datenbanksystem-Migrationen.“	ZFMK, Bonn
VIII	23-24.2.2006	„Perspektiven für den zukünftigen Webauftritt von	Universität Bayreuth

---

GBIF Deutschland.“			
IX	30.11-1.12.2006	„Einsatzmöglichkeiten von <i>DiversityResources</i> zur Verwaltung wissenschaftlicher Bildsammlungen.“	BSM, München
X	14.-15.6.2007	„Datenverwaltung im Rahmen der Forschungsaktivitäten in den DFG-Exploratorien“	Universität Bayreuth
XI	17.-18.12.2007	„Wissensbasierte Verwaltung von Daten und der Einsatz von semantischen Systemen im Rahmen der DFG-Exploratorien“	MPI Jena

---

Im Berichtszeitraum wurden elf mehrtägige Workshops zu unterschiedlichen Themen abgehalten. Auf zwei Workshops (II und VIII) wurden Themen behandelt, die GBIF-Deutschland direkt betrafen, nämlich Datenflüsse im GBIF-Deutschland Netz und der Webauftritt von GBIF-Deutschland (<http://www.gbif.de/>). Zwei weitere Workshops (III und IX) wurden für den Informationsaustausch und die direkte Zusammenarbeit von Softwareentwicklern genutzt. Die beiden 2007 durchgeführten Workshops (X und XI) wurden auf Anfrage aus dem DFG-Großprojekt „Biodiversitätsexploratorien“ veranstaltet und dienten zur Evaluierung der Aufgaben und Lösungsmöglichkeiten für die Datenverwaltung und –bereitstellung in diesem Projekt.

Die Workshops IV-VII befassten sich mit Themen, die das Datenmanagement in BIOTA AFRICA und die mögliche Datenbereitstellung für GBIF betrafen. Dabei wurden neben einer Bestandsaufnahme und der Diskussion der bisherigen Datenstrukturen in BIOTA AFRICA die weitere Vorgehensweise im Datenmanagement sowie die Anbindungsmöglichkeiten an GBIF unter Einbeziehung möglichst vieler IT-Verantwortlicher aus den BIOTA AFRICA Teilprojekten diskutiert. Die Ausgangssituation in BIOTA AFRICA ist gekennzeichnet durch eine starke Heterogenität des Datenpools durch viele Teildisziplinen. Gleichzeitig besteht jedoch die Verpflichtung zur nachhaltigen Bereitstellung der Primärdaten. Es wurden einige Punkte zur Erreichung dieses Ziels (nachhaltige Verfügbarkeit der Primärdaten) auf den Workshops IV-VII formuliert:

- Datenstrukturen vereinheitlichen und internationalen Standards entsprechend organisieren (V, VI, VII)
- CIDOC-CRM als grundlegendes Konzept (IV, V, VI)
- Evaluierung verschiedener standardisierter Datenaustauschformate, um den Datenaustausch gewährleisten zu können (V)
- Zusammenarbeit mit anderen Initiativen wie GBIF, CIDOC-CRM, Diversity-Workbench oder SysTax (IV, V, VI, VII)
- Gemeinsames Glossary für die am betroffenen Verbundprojekt beteiligten Teildisziplinen (VII)



- Weitere Mitarbeiterstelle(n) für Umsetzungsarbeiten und Serveradministration (VII)

Zu den konkreten Ergebnissen der Workshops für das Datenmanagement in BIOTA AFRICA gehört die Entwicklung eines Datenbankmoduls „DiversityEcology“ als mögliche Struktur zumindest für biologische Beobachtungsdaten (Link zum Datenmodell: [http://132.180.63.25/GBIF-ITFG/WS06\\_DiversityEcology\\_ER.pdf](http://132.180.63.25/GBIF-ITFG/WS06_DiversityEcology_ER.pdf)). Des Weiteren wurden die Möglichkeiten und Grenzen des ABCD-Standards für BIOTA AFRICA Forschungsdaten evaluiert. Informationen über Datenstrukturen und Austauschstandards bei anderen Vegetationsdatenbanken (Link zum XML-Austauschstandard ES-Veg: <http://www.floraweb.de/vegetweb/VegetWeb.xsd>) wurden eingeholt.

### 3.2. Bilanz

Die Workshops dienten der Zusammenarbeit verschiedener Arbeitsgruppen und konnten zu einer Entscheidungsfindung bei unterschiedlichen Fragen beitragen. Doch auch Workshops, die vor allem der Vorstellung verschiedener Systeme und dem Informationsaustausch dienten, waren positiv zu bewerten, da gerade die technischen Mitarbeiter seltener die Möglichkeit zum Austausch haben und so die Kommunikation zwischen Softwareentwicklern direkt gefördert werden konnte.

Auf einigen der durchgeführten Workshops konnten Datenbankentwicklungen, wie sie in Deutschland im Bereich Biodiversitäts- und Umweltforschung in den letzten Jahren entstanden sind, analysiert werden. Dabei zeigten sich einige grundsätzliche Probleme bzw. Defizite, auf die im Folgenden genauer eingegangen werden soll.

Beim Beginn von Projekten im Bereich der Biodiversitätsforschung stehen beteiligte Institutionen bzw. Arbeitsgruppen häufig vor der Frage, ob (a) bereits vorhandene Datenhaltungssysteme und damit operierende Applikationen genutzt und den neuen Fragestellungen angepasst, (b) neue Datenbanken und Anwendungen entwickelt, oder aber (c) andernorts bereits existierende Systeme in das eigene integriert werden sollten.

Speziell bei kleineren, stärker spezialisierten Arbeitsgruppen sind die vorhandenen IT-Strukturen oft Einzelplatzlösungen, womit anfallende Biodiversitäts- und Umweltdaten zwar strukturiert, aber nicht durchgängig standardisiert vorgehalten werden. Dies geschieht z. B. anhand von Tabellenkalkulationsprogrammen oder gar bloßen Textdateien, die oft nur nach den Vorstellungen einzelner Projektmitarbeiter (Diplomanden, Doktoranden) entwickelt wurden. Zumeist fehlen auch die technischen Möglichkeiten, um Intranet- oder Internet-basiert Eingaben in verteilte oder zentrale Datenbanken vornehmen zu können.

Jedoch erzeugen auch auf speziellere Fragestellungen fokussierte Arbeitsgruppen nicht nur Daten einer einzigen Datendomaine (z. B. georeferenzierte Beobachtungs-

daten, Messreihen-Daten, pflanzensoziologische Daten) sondern eine Kombination derselben. Diese werden oft gemeinsam in mehr oder weniger ausreichend dokumentierten Datenhaltesystemen eingepflegt und höchst selten in nach Datendomainen eigenständigen, bzgl. ihres Aufbaus und der verwendeten Ontologien gut dokumentierten Datenbanken verwaltet. Die verwendeten Daten(-typen) folgen nur selten einem von internationalen Gremien empfohlenen Standard und die in Einzelprojekten anfallenden Daten sind daher in den seltensten Fällen problemlos miteinander kombinierbar bzw. korrelierbar.

Im Falle eines tatsächlich stattfindendem Datenaustausches zwischen einzelnen Arbeitsgruppen im Rahmen angestrebter gemeinsamer Publikationsarbeit findet dieser zumeist auf traditionellem Wege statt, d. h. nicht automatisiert zwischen interoperablen Datenbanken oder via Webservices.

Unter den geschilderten Bedingungen, d. h. bei Fehlen ausreichender Datenstrukturierung, -standardisierung und -vernetzung, besteht für die langfristige Datenhaltung, -pflege und -bereitstellung das Problem, dass die lokale inhaltliche und technische Pflege oft nicht ohne Weiteres über Jahre hinweg gewährleistet werden kann und speziell die Übergabe von Daten an neue Mitarbeiter problematisch wird. Erschwerend kommt bei Nichtbeachtung internationaler Datenstandards (soweit vorhanden) das Problem einer mangelnden Möglichkeit der Datenverschneidung hinzu.

Es zeigte sich, dass, obwohl meist verschiedene Datenbanken verwendet werden, die Verwendung von standardisierten Datenaustauschformaten in der Regel nicht ausreichend realisiert ist. Des Weiteren wurde deutlich, dass die Modularisierung der Datenhaltesysteme nach Datendomainen in meist unzureichendem Maße gegeben ist, obwohl eine solche den flexibleren Umgang auch mit speziellen Datenbeständen ermöglicht. Eine nachhaltige Datensicherung, Dateneingabe und -pflege erwies sich in verteilten Systemen als ebenso möglich wie bei Zentralhaltung. Aufgrund der Möglichkeit der Vergabe abgestufter Zugriffsrechte an verschiedene Autoritäten erschien die dezentrale Datenhaltung in verteilten Systemen für Projektverbünde sogar deutlich geeigneter.

## **4. Dokumentation der IT-Entwicklungen**

### **4.1. Einleitung**

Zur Dokumentation von IT-Entwicklungen im Rahmen der Biodiversitätsforschung in Deutschland sollten sowohl Verantwortliche als auch programmieretechnisch geschulte Mitarbeiter verschiedenster Projekte um detaillierte Auskünfte auf technischer Ebene gebeten werden. Dazu wurde zunächst eine Checkliste („Merkmalsliste“) bestehend aus Abfragen zu den technischen Spezifikationen von IT-Entwicklungen so-

wie zur Datenhaltung und zu technischen Strukturen an den jeweiligen Institutionen (siehe Anlage 4) erstellt. Unter Verwendung des Datenbanksystems *DiversityDescriptions* in der Version 1.9 beta 8 konnten HTML-Formulare („IT-Entwicklung“ und „Institution“) erstellt werden. Diese wurden auf der Website der IT-Fachgruppe verfügbar gemacht und konnten online über einen Webbrowser ausgefüllt werden. Nach dem Absenden des Formulars konnten die eingegebenen Daten über eine CGI-Schnittstelle wieder in die Datenbank eingelesen werden. Die eingegebenen Daten wurden in HTML-Formularen exportiert, welche dann über die Website passwortgeschützt für die jeweilige Institution zur Verfügung gestellt wurden, und so eingesehen, ergänzt oder aktualisiert werden konnten.

Im März 2004 wurden alle IT-Verantwortlichen und Teilprojektleiter im Rahmen von GBIF-Deutschland per Email über die Aufgaben und Ziele der IT-Fachgruppe informiert und gebeten, Daten zu eigenen IT-Entwicklungen über die Online-Formulare einzugeben. Im Vorlauf zum ersten Workshop der IT-Fachgruppe in München zum Thema „Sammlungsdatenbanksysteme“ wurden die eingeladenen IT-Verantwortlichen gebeten die technischen Spezifikationen zu den jeweiligen Entwicklungen anzugeben. Wegen des sehr geringen Rücklaufs wurde die Dokumentation während der Projektlaufzeit mehrmals durch Telefoninterviews mit einer reduzierten Merkmalsliste ergänzt.

Eine Liste der aufgenommenen IT-Entwicklungen ist in Anlage 5 zu finden.

## **4.2. Zusammenschau der Ergebnisse**

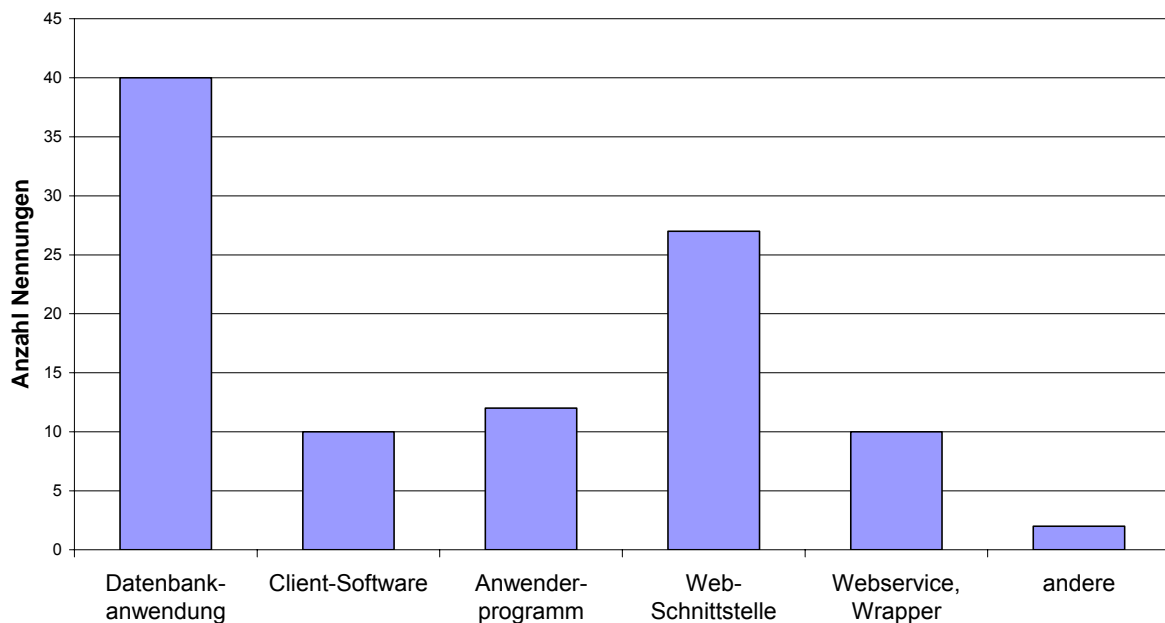
Es konnten 47 IT-Entwicklungen aus dem Bereich der Biodiversitätsforschung mit einer meist ausführlichen technischen Beschreibung in den Bericht aufgenommen werden. Fast alle Angaben wurden im ersten Halbjahr 2007 noch einmal aktualisiert, so dass im Folgenden eine weitgehend aktuelle Übersicht der Entwicklungen gegeben werden kann.

Da beim Ausfüllen der HTML-Formulare nicht zwingend vorgegeben war, welche oder wie viele Fragen beantwortet werden mussten, besitzen die Angaben sehr unterschiedlichen Umfang. Einige der Einträge resultieren beispielsweise aus Telefoninterviews mit einem sehr stark reduzierten Fragenkatalog, die später von den Verantwortlichen nicht mehr ergänzt wurden. Dies erschwert eine gemeinsame Auswertung der Angaben, doch einige Interpretationen sind dennoch möglich. Vor allem die Zusammenstellung der verwendeten Technologien (ohne Anspruch auf Vollständigkeit) mit der zugehörigen Liste der verantwortlichen Entwickler kann auch noch nach Ablauf des Projektes dem Informations- und Erfahrungsaustausch dienen.

### Eckdaten zum IT-Produkt

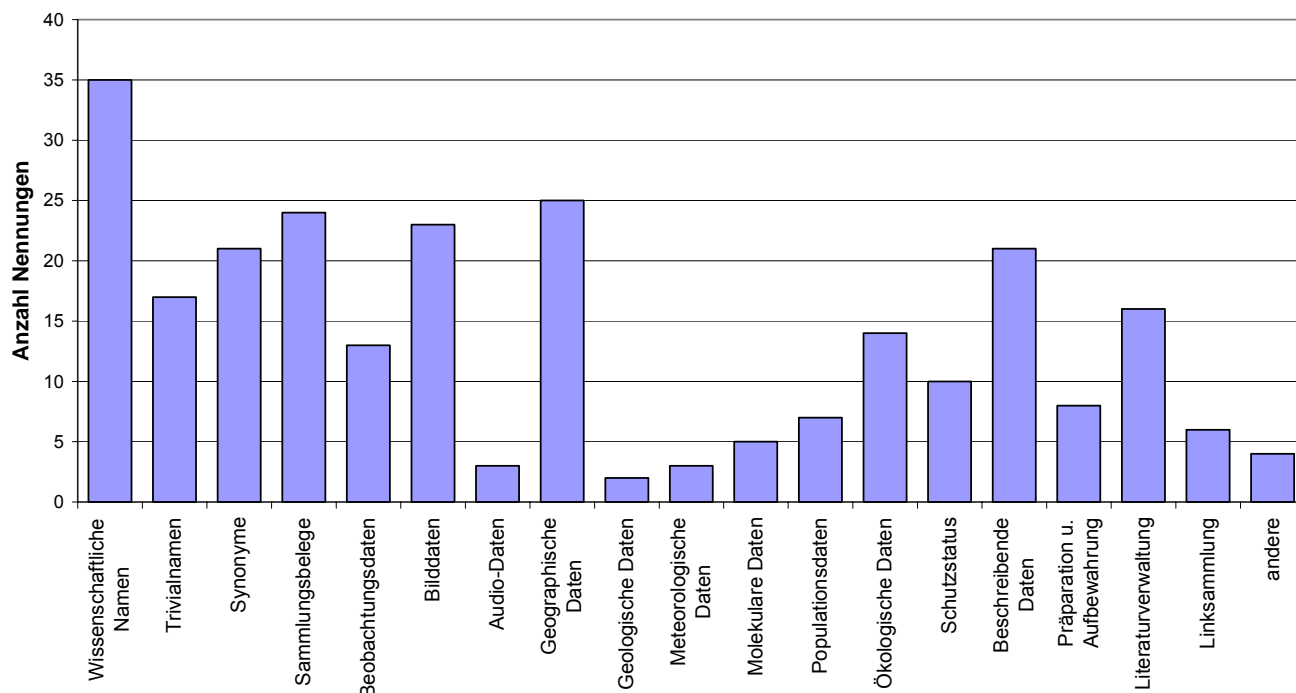
Bei der Frage nach der Art der IT-Entwicklung wurde bei 47 aufgenommenen IT-Entwicklungen 40mal Datenbankanwendung genannt, zehnmal Client-Software, zwölfmal Anwenderprogramm, 27mal Web-Schnittstelle, zehnmal Webservice oder Wrapper und einmal eine andere Art der IT-Entwicklung. Die offensichtlichen Mehrfachnennungen machen deutlich dass bei der Eingabe der Daten oft nicht zwischen Datenbankanwendung und z. B. zugehöriger Webschnittstelle getrennt wurde.

#### Art der IT-Entwicklung



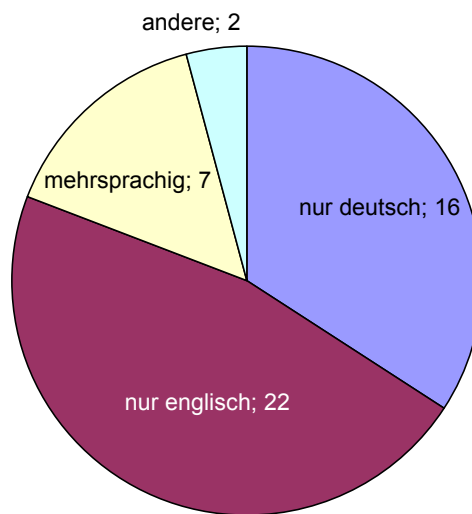
Bei der Frage nach der Art der verwalteten oder analysierten Biodiversitätsdaten wurden am häufigsten wissenschaftliche Namen, geographische Daten, Sammlungsbelege und Bilddaten genannt. Dagegen spielen bei den hier aufgenommenen IT-Entwicklungen Audio-Dateien, geologische Daten, meteorologische Daten und molekulare Daten eine sehr geringe Rolle.

## Art der zu verwaltenden Daten



Die Nutzung der IT-Produkte verläuft zu nahezu gleichen Teilen über eine Client-Server Anbindung (36% der Angaben) oder lokal (40%). Ein Viertel der IT-Entwicklungen wird ausschließlich über das Internet genutzt. Bei der Frage nach der Modularisierung des IT-Produktes wurden insgesamt 40 Angaben gemacht, wonach genau 50% der IT-Entwicklungen aus unabhängigen Modulen bestehen. Bei den eingesetzten http-Servern handelt es sich zu 70% um Apache und zu 25% um Microsoft IIS. Tomcat und der im RAD-Tool 4D angebotene Webserver wurden zusätzlich genannt. Die in der Benutzeroberfläche verwendete Sprache war vor allem Englisch, zu einem etwas geringeren Prozentsatz Deutsch, in einigen Fällen auch mehrsprachig (dann hauptsächlich Englisch und Deutsch).

### In der Benutzeroberfläche verwendete Sprache



Bei 60% der Entwicklungen ist ein Hilfesystem verfügbar (z.T. als online- Hilfe, 19 Anwendungen, z.T. offline, 9 Anwendungen) und in sechs Fällen befand sich das Hilfesystem noch im Aufbau. Es wurde 13mal angegeben, dass kein Hilfesystem existiert. Eine technische Dokumentation ist nur bei 50% der erfassten Entwicklungen vorhanden. In sieben Fällen ist sie im Entstehen.

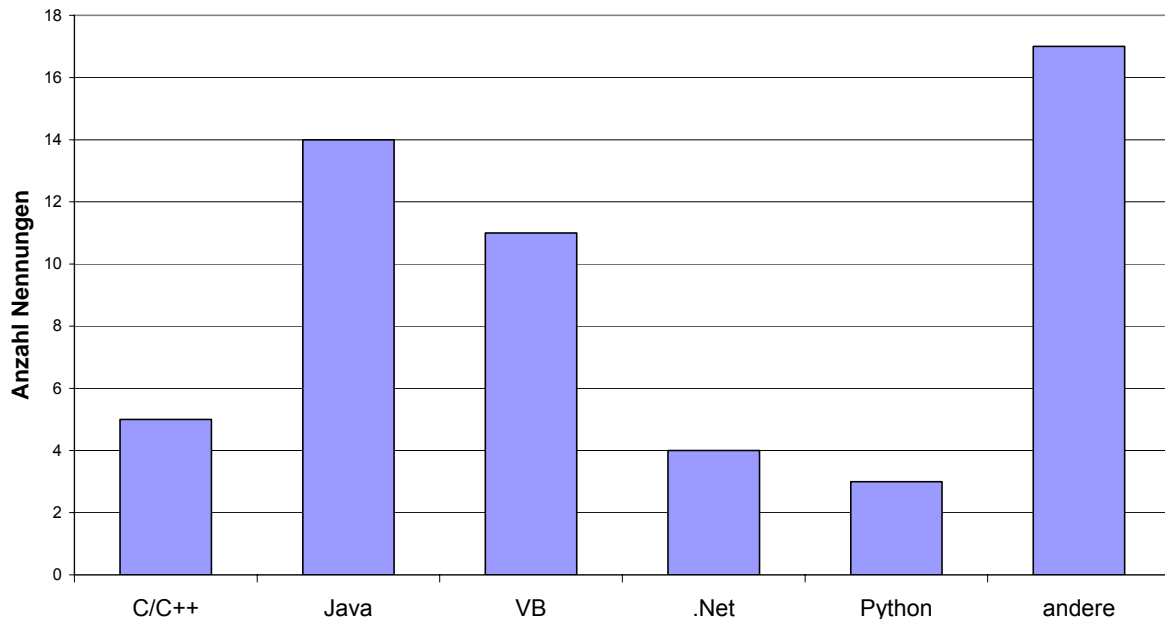
#### **Verfügbarkeit und Lizenzen**

Zu 47 IT-Entwicklungen wurden Angaben zur Verfügbarkeit gemacht, wonach 68% für Dritte verfügbar sind. Zur Lizenzpolitik wurden insgesamt 37 Angaben gemacht. In sieben Fällen ist sie bislang ungeklärt, elf IT-Produkte werden unter OSI-zertifizierte Lizenzen gestellt; dabei wird siebenmal GNU Public License (GPL) und viermal Mozilla Public License (MPL) angegeben. Sieben Produkte sind laut Angaben nicht für Dritte verfügbar, wodurch sich die Frage der Lizenz nicht stellt. Als „andere“ Lizenzen wurde auf eine Oracle Lizenz hingewiesen. In zwei Fällen wird die Frage der Lizenz individuell entschieden.

#### **Programmiersprache und Quelltext**

Bei den zur Auswahl stehenden Programmiersprachen spielte Java mit 26% der Nennungen die wichtigste Rolle, doch 17 Angaben verwandter Programmiersprachen zeigen wie hoch hier die Vielfalt ist. Als zusätzliche Programmiersprachen wurden genannt: Oracle-Tools, Oracle-Forms, Oracle Application Express Release 2.2 (PL/SQL), Oracle-PL/SQL, Delphi, ColdFusion MX, Perl, C, PHP, Visual FoxPro, Java Script, Pascal.

## Verwendete Programmiersprachen



Es wurden nicht für alle IT-Produkte Angaben zu den Quelltexten gemacht, doch von den vorhandenen Angaben kann festgehalten werden, dass für 65% der IT-Produkte die Quelltexte verfügbar sind (insg. 37 Angaben), 76% der Quelltexte kommentiert sind (insg. 33 Angaben) und diese Kommentare zu 53% auf englisch, zu 44% auf deutsch und in einem Fall englisch und deutsch sind (insg. 32 Angaben).

### Datenbankentwicklungen

Es handelte sich bei insg. 40 IT-Produkten um Datenbankentwicklungen. Bei der Frage nach dem verwendeten Datenbankmanagementsystem wurden am häufigsten MS Access (24%), MS SQL Server (24%) und Oracle (22%) genannt. Vier bzw. fünf Nennungen wurden für MySQL bzw. PostgreSQL gemacht. Filemaker, 4D und Visual FoxPro wurden jeweils einmal als Datenbankverwaltungssystem angegeben. Bei der Frage, ob ein publiziertes Datenbankmodell verwendet wurde, wurden 32 Angaben gemacht. In 47% existierte kein publiziertes Modell, in 44% wurde angegeben, dass ein publiziertes Modell verwendet wurde. In 10% der Fälle ist ein Modell in Bearbeitung. In vielen Fällen ist ein ER-Diagramm intern oder über die Website der jeweiligen Entwicklungen verfügbar. Als publizierte Modelle bzw. deren Quelle werden genannt:

- ZOOLIS, BOTIS (Zoologische und Botanische Funddaten)
- Berlin Modell (s. <http://www.bgbm.org/biodivinf/docs/bgbm-model/>) (Taxonomischer Webeditor für Berlin Modell Datenbanken, Interne MS Access Entwicklung Universität Bonn)
- Im Internet publiziertes ER-Diagramm, Tabellenstruktur (SysTax)

- <http://www.diversityworkbench.net> (NaviKey, DiversityCollection, DiversityTaxonNames, DiversityExiccatae)
- SPELDA, J.; ROSENBERG, J.; VOIGTLÄNDER, K. (2003a): The German Myriapod Literature Project (GerMyLit). – African Invertebrates, 44 (1): 325-330. Pietermaritzburg. (JsLit Version8)
- publiziert durch Species2000 (SPICE-Wrapper)

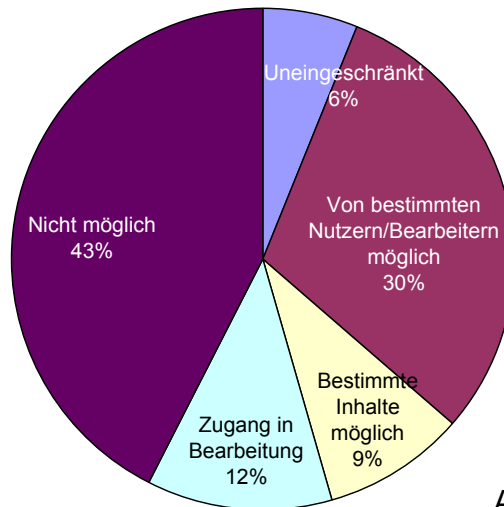
Als Datenbank-Import-Formate wurde neben XML (38% der insg. 34 Angaben) Feldbegrenzer-limitierter Text (44% der Angaben) auch die Möglichkeit von Dateiimporten über ein spezielles Word-Format genannt. Wurde die Option „andere“ gewählt, wurden in den meisten Fällen keine konkreten Angaben gemacht. Als Datenbank-Export-Formate wurde neben XML (33% der insg. 40 Nennungen) Feldbegrenzer-limitierter Text (50% der Nennungen) auch einmal MS Access genannt. Wurde die Option „andere“ gewählt, wurden in den meisten Fällen keine konkreten Angaben gemacht. Der Datenbank-Content-Transfer erfolgt bei den meisten Datenbankentwicklungen über Import/Export-Dateien (59% von insg. 34 Angaben). In acht Fällen wird ODBC („Open Database Connectivity“) und in zwei Fällen JDBC („Java Database Connectivity“) verwendet. Als weitere Möglichkeiten wurde „ActiveX Data Objects“ (ADO), MS-OLEDB und „Borland Database Engine“ (BDE) genannt.

### **Datenbanken und Web**

Eine Datenbankabfrage über das Internet wird in 28 Fällen als möglich angegeben, viermal wurde nur statische Dokumentgenerierung angegeben, 24mal dynamische Dokumentgenerierung. Doch auch hier kamen dreimal Mehrfachnennungen, also beide Möglichkeiten vor. Fünfmal wurde angegeben, dass keine Datenbankabfrage über das Internet möglich ist, zweimal wurde die Option „in Bearbeitung“ gewählt.



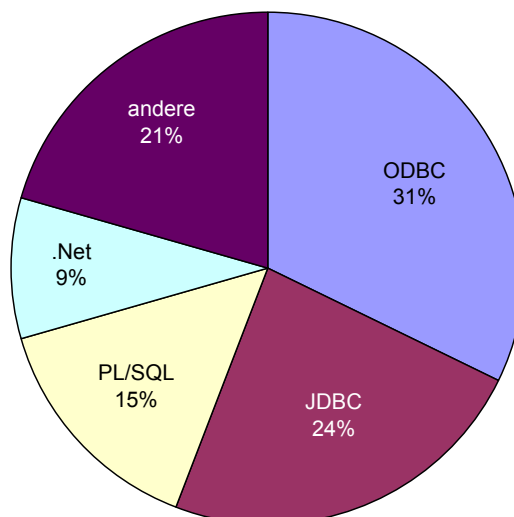
### Editierung von Datenbankinhalten über das Internet



Anzahl Nennungen: 33

Bei einem Großteil der Datenbankanwendungen ist eine Bearbeitung von Datenbankinhalten über das Internet nicht möglich (42% der Nennungen). Bei den Entwicklungen, bei denen ein solcher Zugriff möglich ist, bestand vor allem eine Einschränkung hinsichtlich der Nutzer bzw. Bearbeiter.

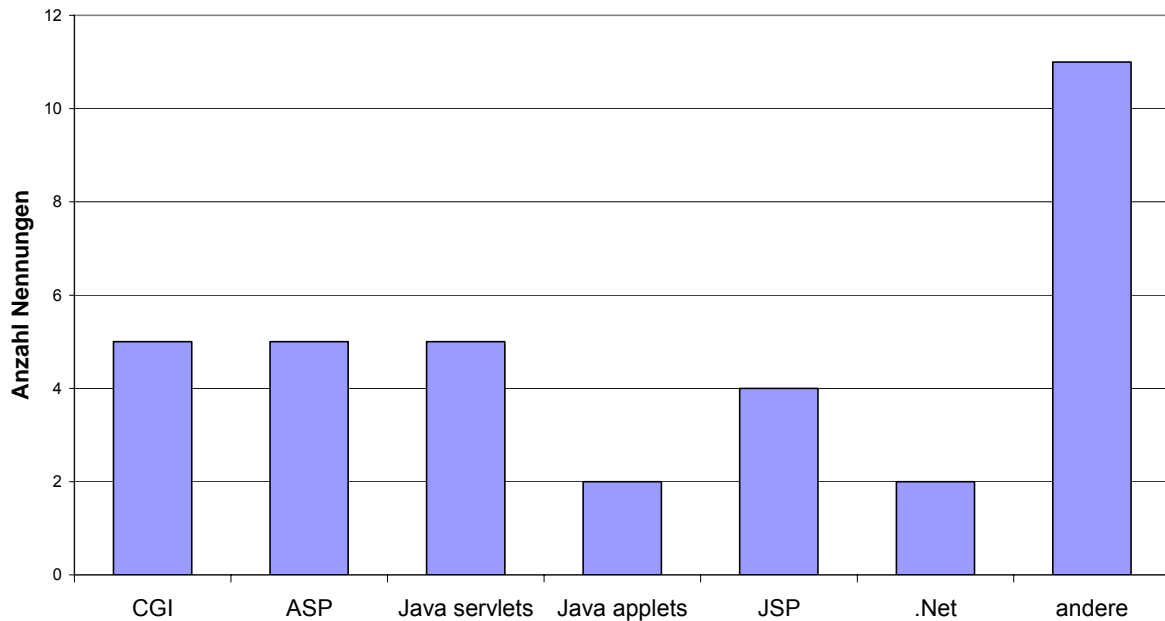
### Schnittstelle zur Datenbank



Anzahl Nennungen: 34

Der Zugriff auf die Datenbanken läuft in den meisten Fällen über ODBC oder JDBC (gemeinsam 55%). Auch .Net und PL/SQL werden angegeben. Als weitere Möglichkeiten wurde „ActiveX Data Objects“ (ADO) und „Borland Database Engine“ (BDE) genannt.

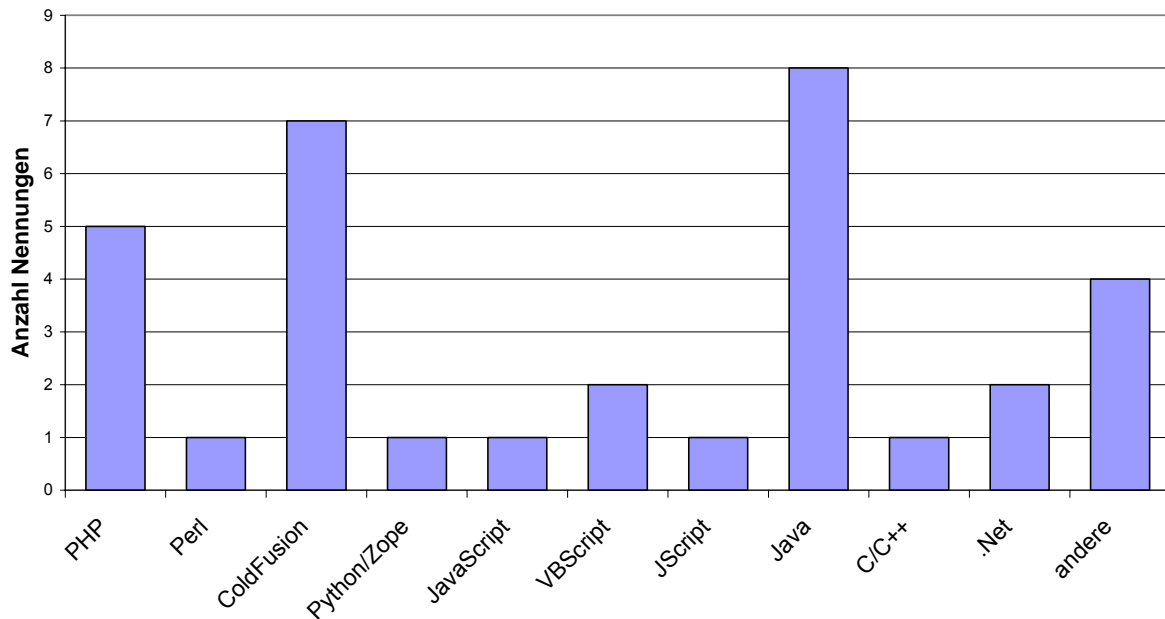
### Technologie der Web-Schnittstelle



Zur verwendeten Technologie der Web-Schnittstelle wurden insg. 34 Angaben gemacht. Danach spielen CGI, ASP und Java Servlets zu gleichen Teilen eine wichtige Rolle. Auch Java Applets, JSP und .Net werden verwendet. Unter der Option „andere“ wurde sechsmal ColdFusion genannt und spielt damit als Webschnittstellen-Technologie die bedeutendste Rolle bei den hier dokumentierten Installationen. Des Weiteren wurde Oracle Application Express (Release 2.2) und die Nutzung der Web-schnittstelle von 4D genannt.

Bei den verwendeten Programmiersprachen für die Web-Schnittstellen haben Java und ColdFusion die größte Bedeutung. Bei der Option „andere“ wird nur einmal Delphi genannt.

### Programmiersprache für Datenbank-Web Schnittstelle



#### 4.3. Interpretation der Ergebnisse

Die Auswertung der hier dokumentierten IT-Entwicklungen zeigt, dass ein breites Spektrum an Technologien verwendet wird, welche die Gesamtheit der in den biologischen Wissenschaften benötigten Datendomänen abdecken. Die betrifft sowohl den Datenbank-Bereich („Container“) als auch Werkzeuge zur Transformation und Übertragung von Daten. Diese Werkzeuge sind jedoch aus vornehmlich historischen Gründen bezüglich ihrer Implementierung sehr divers. Positiv zu bewerten ist auch das Konzept der Modularisierung von Datendomänen, das in einigen Entwicklungen verfolgt wird.

Der größte Prozentsatz der dokumentierten IT-Entwicklungen verwendet aktuelle Technologien (z. B. .Net, Java). Eine Vielzahl der Datentransfer-Protokolle ist web-basiert, hier werden v. a. ODBC und JDBC verwendet.

Im Bereich der Biodiversitätsinformatik ist sicherlich genügend technisches Know-How sowie genügend Implementierungen vorhanden. Die weitere Entwicklung ist jedoch davon abhängig inwieweit sich Projektträger finden, bzw. die öffentliche Hand bereit ist v. a. Mittel zur personellen Unterstützung bereitzustellen.

Die IT-Fachgruppe sieht sich in der Lage derartige strukturelle und technische Weiterentwicklungen auch weiterhin zu begleiten.

## 5. Website der IT-Fachgruppe

Die Website der IT-Fachgruppe ([http://132.180.63.25/GBIF-ITFG/ITFG\\_welcome.html](http://132.180.63.25/GBIF-ITFG/ITFG_welcome.html)) dient zur Information über Ziele und Aufgaben der IT-Fachgruppe und zur Hinterlegung von Workshop-Protokollen. Neben den Protokollen wurden in vielen Fällen auch die Beiträge im pdf-Formate oder Linklisten die interessante Informationen zum Thema des jeweiligen Workshops enthielten zugänglich gemacht.

Die HTML-Formulare für die Angaben zu IT-Entwicklungen sind ebenso auf der Website verfügbar wie die bereits ausgefüllten Formulare, die von den jeweiligen Institutionen passwortgeschützt eingesehen und aktualisiert werden können. Ein Personenverzeichnis der IT-Fachgruppe listet alle Workshopteilnehmer und deren Kontaktdaten auf.



Informationen zum Aufbau des Deutschen Knotens der Global Biodiversity Information Facility

<b>GBIF-D IT-FACHGRUPPE</b>	<a href="#">Home</a> >> <a href="#">Koordination</a> >> IT-Fachgruppe
<b>GBIF-DEUTSCHLAND</b>	<div style="text-align: center;"> <h3 style="color: #8B4513;">Willkommen bei der IT-Fachgruppe!</h3> <p>Die IT-Fachgruppe ist Teil der Koordinationsstruktur von GBIF-Deutschland. Ihre Aufgabe besteht sowohl in der Zusammenführung aller verfügbaren technischen Informationen über existierende Software- bzw. Datenbankentwicklungen (siehe <a href="#">Dokumentation</a>) als auch in der Koordination des Informationsaustausches der im Bereich der Biodiversitätsinformatik aktiven Softwareentwickler im Rahmen von <a href="#">Workshops</a>.</p> <hr/> <p><b><a href="#">Personenverzeichnis der IT-Fachgruppe</a></b></p> <p><b>Ansprechpartner:</b></p> <p>Dr. Alexandra Kehl, Universität Bayreuth (Koordinationsassistentz)</p> <p>Anton Güntsch, Botanisches Museum und Botanischer Garten, Berlin-Dahlem (Sprecher der IT-Kommission)</p> <p>Prof. Dr. Gerhard Rambold, Universität Bayreuth (Koordination)</p> <hr/> <p><b>Workshops der IT-Fachgruppe:</b></p> <p>Auf den Workshops soll den eingeladenen IT-Fachleuten bzw. Softwareentwicklern Gelegenheit geboten werden, technische Details ihrer IT-Entwicklungen vorzustellen und die anderer Projekte kennenzulernen. Dabei sollen sowohl überlappende Aktivitäten und Redundanzen aufgedeckt als auch die Entwicklung gemeinsamer plattformunabhängig einsetzbarer Anwendungen oder Webschnittstellen abgestimmt werden.</p> </div>
<b>GBIF-INTERNATIONAL</b>	<a href="#">Home</a> <a href="#">Deutsche Delegation</a> <a href="#">CIRCA Archiv</a>

## 6. Danksagung

Wir danken Prof. Dr. S. Jablonski und B. Volz (Lehrstuhl für Angewandte Informatik IV, Universität Bayreuth) für hilfreiche Korrekturen am Manuskript. Für die organisatorische Unterstützung bei der Ausrichtung der Workshops bedanken wir uns herzlich bei Dr. Dagmar Triebel und Ingrid Sebek (Botanische Staatssammlung München), Anton Güntsch (BGBM Berlin), Dr. Christoph Schultz und Dr. Melanie Vogel (Universität Würzburg), Lothar Menner (FIS, Frankfurt a. M.), Dr. Karl-Heinz Lampe (ZFMK, Bonn) und Dr. Jens Nieschulze (MPI BGC, Jena).

## 7. Anlagen

### Anlage 1: Protokolle zu den Workshops

Nr.	Datum	Thema	Seite
I	23.-24.4.2004	„Sammlungsdatenbanksysteme“	19
II	28.-29.10.2004	„Spezifikationen zum Datentransfer und zukünftige Datenflußszenarien.“	30
III	6.-7.1.2005	„Konzept für eine Rich Client Plattform zur Integration von <i>DiversityWorkbench</i> Komponenten.“	39
IV	12.-13.5.2005	„Integration und Anbindung von BIOTA-AFRICA Forschungsdaten in das GBIF-Netzwerk.“	46
V	22.-23.6.2005	„Erstellung eines Datenmodells für die in BIOTA erfassten Standorteigenschaften (Klimatologie, Bodenkunde, Fernerkundung).“	51
VI	14.-15.7.2005	„Datenbanken zur Verwaltung ökologischer Daten in BIOTA und GBIF am Beispiel botanischer und zoologischer Beobachtungsdaten: Anforderungen, Gemeinsamkeiten, Perspektiven.“	56
VII	29.-30.9.2005	„Für den Einsatz in BIOTA AFRICA geeignete Datenbanksysteme und Clients; Notwendigkeit BIOTA-spezifischer Anpassungen; Möglichkeiten bzw. Strategien für Datenbanksystem-Migrationen.“	62
VIII	23-24.2.2006	„Perspektiven für den zukünftigen Webauftritt von GBIF Deutschland.“	65
IX	30.11-1.12.2006	„Einsatzmöglichkeiten von <i>DiversityResources</i> zur Verwaltung wissenschaftlicher Bildsammlungen.“	76
X	14.-15.6.2007	„Datenverwaltung im Rahmen der Forschungsaktivitäten in den DFG-Exploratorien“	83

XI	17.-18.12.2007	„Wissensbasierte Verwaltung von Daten und der Einsatz von semantischen Systemen im Rahmen der DFG-Exploratorien“	95
----	----------------	--	----

## **Protokoll zum Workshop I der IT-Fachgruppe (GBIF-D) „Sammlungsdatenbanksysteme“**

Beginn:	23. April 2004	14.00 Uhr
Ende:	24. April 2004	16.00 Uhr
Ort:	Botanische Staatssammlung München bzw. LMU Botanik, Kurssaal 2 Menzinger Straße 67, 80638 München	
Leitung:	Gerhard Rambold	
Protokoll:	Alexandra Kehl	

Teilnehmer: Helmut Franz (Nationalpark Berchtesgaden, *Datenbank des Nationalpark Berchtesgaden*), Thorsten Ludwig (Universität Ulm, *SysTax*), Lothar Menner (Forschungsinstitut Senckenberg, Frankfurt a. M., *SeSam*), Gerhard Muche (Universität Hamburg, *BioInput*), Andreas Schnock (Naturkundemuseum Berlin, *BIODAT*), Stephanie Sobek (BGBM Berlin, *Specify*), Dirk Striebing (Naturkundemuseum Berlin, *BIODAT*), Markus Weiss (BSM München, *DiversityCollection*)

### **Hintergrund und Ziele**

Ziel dieses Workshops ist es, eine Übersicht über die verschiedenen Entwicklungen und deren technischen Details und weniger über die Anwendungsmöglichkeiten zu bekommen, sowie Kooperationsmöglichkeiten aufzudecken. Innerhalb der Diversität von Werkzeugen die Ähnliches (aber nicht dasselbe) erfüllen, sollen Stärken und Schwächen (diese aber nur um Änderungsmöglichkeiten herauszuarbeiten) analysiert werden. Zwar wird eine Effizienzsteigerung angestrebt, doch sollte eine gewisse Diversität an IT-Entwicklungen möglichst erhalten werden. Ein Bedarf an Arbeitsteilung ist jedoch nicht zu übersehen.

Für zukünftige Anträge im Bereich Biodiversitätsinformatik sollten auf diesem Workshop Argumente gesammelt werden.

### **Vorbemerkungen**

Am 30.3.2004 wurde an Kontaktpersonen und Projektleiter im Bereich Biodiversitätsinformatik in Deutschland ein Rundmail mit der Bitte zum Ausfüllen eines Online-Formulars zu technischen Details zu IT-Entwicklungen gesendet.

Die Auswahl der Workshopthemen für die sieben geplanten Workshops und insbesondere für den ersten Workshop, sowie die Auswahl der Teilnehmer wurde mit der IT-Kommission (Arbeitstreffen am 17.3.2004 in Stuttgart) abgestimmt.

Die Anzahl der Teilnehmer pro Workshop sollte in Übereinstimmung mit der IT-Kommission auf ca. 10–12 begrenzt sein, daher konnten bei der Einladung nicht alle Entwicklungen berücksichtigt werden.

Die Bereitstellung der Daten aus den Online-Formularen sollte nach Diskussion in der IT-Kommission (Arbeitstreffen am 17.4.2004 in Stuttgart) eher restriktiv gehandhabt werden, dennoch wird für ITFG-Beteiligte eine Online-Abfrage technischer Daten der Projekte eingerichtet werden. Zurzeit besteht Zugang zu den beiden Leerformularen „Institution“ und „IT-Entwicklung“ über ein allen Angeschriebenen bekannt gegebenes passwort-geschütztes Login. Das Formular „IT-Entwicklungen“ wurde von allen Teilnehmern vor dem Workshop über das Internet ausgefüllt und die Einträge z. T. während des Workshops korrigiert. Die Möglichkeit der Korrektur und Ergänzung ausgefüllter Formulare über personalisierte Zugänge besteht, die entsprechenden Passwörter werden den Teilnehmern des Workshops mit dem Protokoll zugesandt.

Vom Vorsitzenden der Knotenkoordinatoren wurde am 5.5.2004 der Kurzbericht über Aktivitäten der ITFG für die Sitzung am 7.5.2004 in Bonn angefordert; darin wurde über die Aktivitäten der IT-Fachgruppe seit Januar 2004 berichtet und ein kurzer Überblick über Teilnehmer, Ablauf und Ergebnisse des ersten Workshops gegeben.

Die Verteilung des vorläufigen Protokolls an die Teilnehmer zur kritischen Durchsicht erfolgte am 14.5.2004.

Die endgültige Fassung des Protokolls wurde dem Sprecher der IT-Kommission (Holstein) zugesandt zur Weiterleitung an die IT-Kommission bzw. den Vorsitzenden der Knotenkoordinatoren.

## **Beiträge**

### **„Übersicht ABCD-Schema“**

(M. Weiss und D. Striebing)

Das ABCD-Schema wurde entwickelt aus dem Bedarf der Harmonisierung zwischen den verschiedenen Datenbanken. Über GBIF werden die Daten von den einzelnen Knoten bzw. Datenbanken über das ABCD Schema abgefragt und dem Benutzer zur Verfügung gestellt. Jedes Projekt bzw. jede Datenbank bleibt Halter der eigenen Daten und kann entscheiden, welche Daten freigegeben werden sollen. Das ABCD Schema muss an die jeweilige Datenbankstruktur angepasst werden.

### **„BIODAT“**

(D. Striebing, Museum Alexander Koenig, Bonn / Museum für Naturkunde, Berlin)



Bedarfsorientierte Entwicklung für den eigenen Kreis im Naturkundemuseum Berlin, bisher nur zoologisch ausgerichtet, soll nun auch für die Abteilung Paläontologie angepasst werden.

Quellcode und Software ist für Dritte nicht verfügbar. Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Sammlungsbelege (Naturgeschichtliche Sammlungen), Bilddaten, geographische Daten und beschreibende Daten. Das IT-Produkt kann über Client-Server Anbindung genutzt werden. Art des IT-Produktes: Datenbankanwendung, Client-Software, Anwenderprogramm und Web-Schnittstelle.

Die Software ist modularisiert; und läuft als stabile Version. Die in der Benutzeroberfläche verwendete Sprache ist Deutsch. Ein Hilfesystem oder eine Benutzeranleitung ist in Bearbeitung. Eine technische Dokumentation ist in Bearbeitung. Programmiersprache(n) des IT-Produkts: C/C++, VB und andere (Pascal). Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar. Verwendetes Datenbank-Verwaltungssystem: MS SQL Server. Ein publiziertes Datenmodell existiert nicht. „Authority files“ wurden nicht integriert. Eine Datenbankabfrage online ist in Bearbeitung. Die Datenbankabfrage online ist eingeschränkt. Die Datenbankinhalte können online nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC oder ADO. Realisierung (Technologie) der Web Schnittstelle: ASP. Programmiersprache für Datenbank-Web Schnittstelle: JavaScript und VBScript.

Kooperationsbedarf:

- ISO-Standard Ländercodes werden häufig geändert; generell wären gemeinsame Standards gewünscht (Idee: Einrichten eines zentralen Servers für Listen in Deutschland);
- Taxonomie-Modul evtl. bei Einbeziehung der Botanik, generell Austauschmöglichkeit mit anderen Softwareentwicklern, die an ähnlichen Problemen arbeiten.

### „BiotInput“

(G. Mucho, Universität Hamburg)

Die Software dient der Eingabe, Aufbereitung, Haltung, Verwaltung und Auswertung von vegetationsökologischen und floristischen Daten. Konstruiert wurden Datenbank und Programm für hauptsächlich vegetationsökologische und floristische Aufnahmen in Afrika (BIOTA AFRICA, Impetus Marokko). Die Datenstrukturen wurden den Benutzeranforderungen entsprechend entwickelt, so dass die im Gelände (bisher) verwendeten Feldbücher und Datenblätter dv-technisch abgebildet werden können. Ein-

gabeln und Hilfsroutinen erleichtern den Benutzern die notwendigen manuellen Eingaben und inhaltlichen Aufbereitungsschritte.

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: andere Lizenz. Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Sammlungsbelege, Beobachtungsdaten (Monitoring, Erfassungen), geographische Daten, ökologische Daten und beschreibende Daten. Das IT-Produkt kann lokal (auch über lokale Netzwerke), d.h. nicht über Client-Server-Anbindung genutzt werden. Art des IT-Produktes: Datenbankanwendung und Anwenderprogramm.

Das IT-Produkt ist modularisiert und läuft als stabile Version (*Release vom 23. Januar 2004*). Die in der Benutzeroberfläche verwendete Sprache ist Englisch. Ein Hilfesystem oder eine Benutzeranleitung ist in Bearbeitung. Eine technische Dokumentation ist in Bearbeitung. Programmiersprache(n) des IT-Produkts: C/C++. Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar. Verwendetes Datenbank-Verwaltungssystem: MS Access (*ab Version 97*). Ein publiziertes Datenmodell existiert. Bezeichnung des Datenmodells: „ER-Diagramm“. Eine Datenbankabfrage online ist nicht möglich.

Kooperationsbedarf:

- Veröffentlichung über http; Erstellung von Benutzerschnittstellen;
- Abschätzung Aufwand und Nutzenaspekte alternativer Datenbanktechnologien; Einbindungsmöglichkeiten von Authority-Files.

### **„Datenbank des Nationalpark Berchtesgaden“**

(H. Franz, Nationalpark Berchtesgaden)

Hauptinhalt sind Luftbildinterpretationen;

Die IT-Entwicklung ist für Dritte nicht verfügbar. Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Beobachtungsdaten (Monitoring, Erfassungen), geographische Daten, meteorologische Daten, ökologische Daten und Schutzstatus. Das IT-Produkt kann über Client-Server Anbindung genutzt werden. Art der Software: Datenbankanwendung (und GIS-Anwendung).

Das IT-Produkt ist modularisiert und läuft als stabile Version. Die in der Benutzeroberfläche verwendete Sprache ist Deutsch. Ein Hilfesystem oder eine Benutzeranleitung ist nicht verfügbar. Eine technische Dokumentation ist nicht verfügbar. Programmiersprache(n) des IT-Produkts: andere (Oracle-Tools). Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar. Verwendetes Datenbank-Verwaltungssystem: Oracle (Version 9.2 R2 + ArcGIS 8.3). Ein publiziertes Datenmodell existiert. Bezeichnung des Datenmodells: ZOO LIS, BOTIS. „Authority files“ wurden nicht integriert. Eine Datenbankabfrage online ist nicht möglich. Die Daten-

bankinhalte können online nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC, JDBC oder JSP.

Es ist keine Migration z. B. nach PostgreSQL beabsichtigt wegen vertraglicher Verpflichtungen und aus Kompatibilitätsgründen mit dem Datenbanksystem des Bayerischen Umweltministeriums (Oracle).

Kooperationsbedarf:

- externes Modul für deskriptive Daten;
- bzgl. Taxonomie evtl. Kooperation mit SysTax.

### „SysTax“

(T. Ludwig, Universität Ulm)

Integriertes Sammlungsdatenbanksystem für Lebend- und Herbarmaterialien. Zentrale Verwaltung der Daten, d.h. Hosting der Daten für verschiedene Institute.

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: andere Lizenz. Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Sammlungsbelege (Naturgeschichtliche Sammlungen), Beobachtungsdaten (Monitoring, Erfassungen), Bilddaten, geographische Daten, Schutzstatus, beschreibende Daten, Literaturverwaltung, Linksammlung und andere (Pflanze-Tier-Interaktionen, Multimedia-Daten allgemein, Leihverkehr (Sammlungsverwaltung), Verwaltung Botanischer Gärten (incl. Erstellung des Index Seminum). Das IT-Produkt kann über Client-Server Anbindung oder lokal genutzt werden (SysTax-light als stand-alone-Version mit Export-Schnittstelle). Art des IT-Produktes: Datenbankanwendung, Client-Software, Anwenderprogramm, Web-Schnittstelle und Webservice, Wrapper.

Das IT-Produkt ist nicht modularisiert und läuft als stabile Version (4.1). Die in der Benutzeroberfläche verwendete Sprache ist mehrsprachig (deutsch und englisch). Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar. Eine technische Dokumentation ist nicht verfügbar. Verwendete Programmiersprache(n): Oracle-Forms, Oracle-PL/SQL, Perl, C. Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar. Verwendetes Datenbank-Verwaltungssystem: Oracle (8.1.7.4). Ein publiziertes Datenmodell existiert. Bezeichnung des Datenmodells: „ER-Diagramm“, Tabellenstruktur. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist nicht eingeschränkt (vom Benutzer für die Online-Anzeige nicht freigegebene Daten sind weder anzeigbar noch suchbar). Die Datenbankinhalte können online nur von bestimmten Usern editiert werden (nur durch Verwendung der Client-Software, nicht im WWW). Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): andere.

Realisierung (Technologie) der Web Schnittstelle: CGI. Programmiersprache für Datenbank-Web Schnittstelle: Perl und C/C++.

Kooperationsbedarf:

- GIS-Anbindung;
- evtl. externes Literatur-Modul.

### „Specify“

(S. Sobek, *Specify Software Project, University of Kansas Biodiversity Research Center (Betreuung in DE: BGBM Berlin)*)

Integriertes Sammlungsdatenbanksystem

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: OSI-zertifizierte open source Software. (Weiter-)Entwicklung erfolgt momentan durch ein Entwicklerteam in den USA. Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Sammlungsbelege (Naturgeschichtliche Sammlungen), Beobachtungsdaten (Monitoring, Erfassungen), Bilddaten, geographische Daten, ökologische Daten, Schutzstatus, beschreibende Daten, Literaturverwaltung, Linksammlung und andere. Das IT-Produkt kann über Client-Server Anbindung (Keine Online-Nutzung über das WWW, nur in lokalen Netzwerken) oder lokal genutzt werden. Art des IT-Produktes: Datenbankanwendung, Anwenderprogramm, Web-Schnittstelle und Webservice, Wrapper (DIGIR, BioCASE Wrapper sind in Arbeit).

Das IT-Produkt ist modularisiert (Specify beinhaltet den RAVE Report Designer von Nevrona zum Gestalten von Etiketten und Berichten. Außerdem ist die Nutzung des Programms insofern modular, als dass in viele Tabellen Daten eingegeben werden können, die nicht sofort mit Sammlungsobjekten oder Taxa verknüpft werden müssen (z. B. Literatur, Fundorte, Projekte etc.) und läuft als stabile Version. Die in der Benutzeroberfläche verwendete Sprache ist mehrsprachig (Momentan sind Englisch und Deutsch implementiert, weitere Sprachen sind vorgesehen und können sobald eine Übersetzung vorliegt implementiert werden.). Ein Hilfesystem oder eine Benutzeranleitung ist online und offline verfügbar. Eine technische Dokumentation ist online verfügbar (Vereinfachte Datenmodelle und ein Data Dictionary sind online, weitere technische Dokumentation ist z. Zt. nicht verfügbar, jedoch in Arbeit.) oder in Bearbeitung. Programmiersprache(n) des IT-Produkts: VB und andere (ca. 99.9 % Delphi). Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar. Verwendetes Datenbank-Verwaltungssystem: MS SQL Server. Ein publiziertes Datenmodell existiert bzw. ist in Bearbeitung (Vereinfachtes Datenmodell ist online, komplexeres ist in Arbeit. Bislang wird alte MS Access Version des Datenmodells genutzt). „Authority files“ wurden integriert. Eine Datenbankabfrage online ist mit statischer oder mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist nicht

eingeschränkt (Wird durch Datenhalter bestimmt). Die Datenbankinhalte können online nicht editiert werden (z. Z. noch nicht, aber angedacht). Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC. Realisierung (Technologie) der Web Schnittstelle: CGI. Programmiersprache für Datenbank-Web Schnittstelle: andere (Delphi).

Kooperationsbedarf:

- Darstellung von organismischen Interaktionen (Symbiosen; Wirt-Parasit-Interaktionen)
- Der Bereich Taxonomie ist nur unzureichend abgedeckt und sollte evtl. ausgelagert werden

### „SeSam“

(L. Menner, Forschungsinstitut Senckenberg, Frankfurt a. M.)

Sammlungserfassung, angepasst an Bedürfnisse des Hauses (Forschungsinstitut Senckenberg).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: wird individuell entschieden. Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Sammlungsbelege (Naturgeschichtliche Sammlungen), Bilddaten, geographische Daten, ökologische Daten und Literaturverwaltung. Art des IT-Produktes: Datenbankanwendung und Anwenderprogramm.

Das IT-Produkt ist nicht modularisiert; läuft als stabile Version (1.0). Die in der Benutzeroberfläche verwendeten Sprachen sind Deutsch und Englisch. Ein Hilfesystem oder eine Benutzeranleitung ist in Bearbeitung. Eine technische Dokumentation ist in Bearbeitung. Programmiersprache(n) des IT-Produktes: andere. Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar (wird jedoch individuell entschieden). Verwendetes Datenbank-Verwaltungssystem: MS SQL Server (7). Ein publiziertes Datenmodell existiert nicht. „Authority files“ wurden integriert. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist nicht eingeschränkt. Die Datenbankinhalte können online global editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC (zwischen Datenbank und Webserver). Realisierung (Technologie) der Web Schnittstelle: ASP. Programmiersprache für Datenbank-Web Schnittstelle: VBS und JavaScript.

Der Zugriff auf die Datenbank erfolgt über einen Webbrowser (unabhängig vom Betriebssystem), eine Version („SeSam light“) zur offline Dateneingabe existiert nicht und ist auch nicht in Planung. Dafür könnten Anwendungen wie Specify verwendet werden.

Kooperationsbedarf:

- externes Literatur-Modul (als Webservice?).

### „DiversityCollection“

(M. Weiss, Botanische Staatssammlung München)

Sammlungserfassung mit Schwerpunkt Beziehungen zwischen den Organismen/Taxa abzubilden.

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: OSI-zertifizierte open source Software. Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: Sammlungsbelege (Naturgeschichtliche Sammlungen) und geographische Daten. Das IT-Produkt kann über Client-Server Anbindung genutzt werden. Art des IT-Produktes: Datenbankanwendung und Client-Software.

Das IT-Produkt ist modularisiert (*Diversity Workbench*). Die in der Benutzeroberfläche verwendete Sprache ist deutsch (neue Version mehrsprachig) und englisch. Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar und in Bearbeitung (neue Version). Eine technische Dokumentation ist online verfügbar oder in Bearbeitung (Neue Version). Programmiersprache(n) des IT-Produkts: VBA (alte Version), .Net (C# neue Version). Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar. Verwendetes Datenbank-Verwaltungssystem: MS SQL Server (neue Version) oder MS Access. Ein publiziertes Datenmodell existiert. „Authority files“ wurden integriert. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist eingeschränkt. Die Datenbankinhalte können online nur von bestimmten Usern editiert werden und nur in bestimmten Bereichen editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC. Realisierung (Technologie) der Web Schnittstelle: andere (ColdFusion). Programmiersprache für Datenbank-Web Schnittstelle: ColdFusion.

– ein Modul innerhalb der *Diversity Workbench*

Kooperationsbedarf:

- GIS-Anwendungen, bzw. -Anbindung;
- bisher nur als Datenmodell vorliegendes Taxonomie-Modul der *Diversity Workbench* (*DiversityTaxonomy*) sollte in Kooperation mit anderen Arbeitsgruppen implementiert werden; evtl. auch *DiversityReferences* als Literaturmodul).

## **Verlauf der Diskussion und Ergebnisse**

### **Verfügbarkeit: Software und Quellcode**

Ist bei einigen der diskutierten Entwicklungen (noch) nicht gegeben, wird von allen am Workshop Beteiligten gewünscht. Problematisch ist für viele Institutionen dabei die Gewährleistung des Supports, die aus zeitlichen, personellen und finanziellen Gründen oft nicht möglich ist.

### **Programmiersprachen: Notwendigkeit und Möglichkeiten der Portierung**

Notwendigkeiten der Portierung einer Entwicklung in eine andere Programmiersprache oder Überführung auf eine andere technologische Plattform wurden nicht identifiziert.

### **Harmonisierung: Notwendigkeit bzw. Möglichkeiten von Redesign bezüglich der Feldbenennungen etc.**

Die Notwendigkeit einer Harmonisierung von Felddefinitionen der verschiedenen Datenbanken ist kurzfristig nicht erforderlich. Im Gegenteil, die Installation von Wrappern (nach dem ABCD Schema) erlaubt die Beibehaltung bestehender Strukturen. Eine mittelfristige Anpassung an (gewisse) Elemente des ABCD Schema wird für einige der vorgestellten Entwicklungen dennoch nicht ausgeschlossen, hat jedoch keine Priorität. Vorteile hierfür könnte man sehen in a) der Bedeutung des „common namespace“; b) Unterstützung anderer, noch nicht etablierter Mechanismen im Zusammenhang mit remote Abfragen verteilter Datenbanken.

### **Datenmodell/Modularisierung: Notwendigkeit und Möglichkeiten von Redesign und partieller Anwendung**

Für einige Datenbanksysteme sollte erwogen werden, vorgesehene Strukturen zur Speicherung von taxonomischen Daten oder von Literatur nicht weiter auszubauen und stattdessen eine Integration bestehender Module der Datenbank-Suite *Diversity Workbench* in Erwägung zu ziehen (*DiversityTaxonomy*, *DiversityReferences*). Die Datenbank des Nationalpark Berchtesgaden könnte diesbezüglich mit *SysTax* kooperieren).

### **Interoperabilität und Datenaustauschformate: ABCD Schema etc.**

Wird ein Thema im Rahmen des nächsten Workshops sein.

### **Charakterisierung und Einordnung in die Diversität bestehender Datenbank-Lösungen**

Die meisten der Datenbank-Entwicklungen dienen in erster Linie der Erfassung von Lebend- und Totmaterial-Sammlungen. Während es sich bei *SysTax* und *Specify* um integrierte Lösungen handelt, sind *BIODAT*, *BiotaInput* und *SeSam* unmittelbar dem

Bedarf der Sammlung angepasste Systeme. *DiversityCollection* ist dagegen Teil einer modularisierten Datenbank-Suite (*Diversity Workbench*).

Zu Details: siehe unter den Einzeldarstellungen.

### **Kooperationsbedarf**

Wird allgemein als notwendig betrachtet, vor allem bezüglich:

- Integration von Modulen für ergänzende Inhalte (hier v.a. Taxonomie, Literatur, geographischer Gazetteer)
- Austausch von Algorithmen und Funktionen als Lösungspool von (allgemeingültigen) Problemen (z. B. Vor/Nachteile von Datumsformaten, Hilfsfelder für hierarchischen Suchabfragen)
- Problematik: Interaktionen zwischen Taxa
- Informationsaustausch zwischen den Softwareentwicklern

### **Argumente zur Erhaltung einer gewissen Diversität an Datenbankentwicklungen**

Wissenschaftler und technische Mitarbeiter (z. B. an Sammlungen) fordern an den konkreten Bedarf angepasste Strukturen. Weniger spezifische Lösungen wirken aus der Wahrnehmung potenzieller Benutzer eher hemmend auf den Arbeitsfluss. Die Erarbeitung von effizienten IT-Anwendungen erfordert die Eingrenzung auf das konkret Wesentliche. Nur so sind Übersichtlichkeit und schnelle Antwortzeiten bei komplexen Mensch-Rechner-Interaktionen, die bei anspruchsvollen Datenbankanwendungen unabdingbar sind, technisch überhaupt möglich. Umfassende Standardsysteme geben diesen Effizienzvorteil notgedrungen auf, um den Anwenderkreis zu vergrößern.

Die Anpassung einer Fremdsoftware an die speziellen Bedürfnisse ist oft aufwändiger als eine eigene Entwicklung.

Jede gute Standardisierung benötigt einen Vorlauf konkurrierender Lösungsansätze für thematisch ähnliche Problemstellungen. Die Systeme zur Erfassung, Verwaltung und Auswertung biologischer Daten befinden sich nach wie vor in der Prä-Standardisierungsphase.

### **Organisatorisches zum Workshop**

#### **Verbesserungsvorschläge für künftige Workshop**

Größe der Gruppe max. 12 Teilnehmer

Bereitstellung von Präsentationsunterlagen



### **Internet-Portal der IT-Fachgruppe**

Es existiert eine öffentliche Homepage der IT-Fachgruppe mit allgemeinen Informationen und den Ankündigungen der Workshops, von der man Zugang zu den Online-Formularen über ein allgemeines Passwort hat. Jede Institution hat einen eigenen passwortgeschützten Zugang zu ihren bereits ausgefüllten Formularen, um diese zu ändern und zu aktualisieren. Teilnehmer der Workshops haben ebenfalls Zugang zu speziellen Informationen. Bei Bedarf wird an der Universität Bayreuth ein CVS installiert.

### **Bereitstellung der Daten und Ergebnisse des Workshops**

Ein ausführliches Protokoll wird den Teilnehmern des Workshops und deren Projektleitern, sowie Herrn Dr. J. Holstein als Sprecher der IT-Kommission zugesandt. In der nächsten Sitzung der IT-Kommission wird ausführlich über den Workshop berichtet werden, Anregungen werden weitergegeben.

## **Protokoll zum Workshop II der IT-Fachgruppe (GBIF-D) „Spezifikationen zum Datentransfer und zukünftige Datenfluss-Szenarien“**

Beginn:	28. Oktober 2004	12.00 Uhr
Ende:	29. Oktober 2004	13.00 Uhr
Ort:	Konferenzraum des Botanischen Museums Berlin, Königin-Luise-Straße 6–8, Berlin-Dahlem	
Leitung:	Gerhard Rambold	
Protokoll:	Alexandra Kehl	
Teilnehmer:	Evelin Boos (Berlin), Anton Günsch (BGBM Berlin), Gregor Hagedorn (BBA Berlin), Andrea Hahn (BGBM Berlin), Andreas Kohlbecker (BBA Berlin), Alexander Kroupa (Science4you, Bayreuth), Ekaterina Langer (FU Berlin), Thorsten Ludwig (Universität Ulm), Gerhard Mücke (BIOTA AFRICA, Universität Hamburg), Clemens Oertel (Universität Tübingen), Klaus Schild (FU Berlin), Lutz Suhrbier (FU Berlin)	

### **Ziele**

Ziel dieses Workshops ist es, einen Überblick über Datenflüsse in den verschiedenen Projekten zu erhalten, und im Weiteren Datenflussszenarien zu entwerfen, die für die nächste Antragsphase für GBIF-D benötigt werden. Diese Datenflüsse oder „Use Cases“ sollen die Voraussetzungen und grundlegenden notwendigen Schritte für den Transfer unterschiedlicher Datentypen zeigen.

### **Vorbemerkungen**

Die IT-Fachgruppe ist Teil der Organisationsstruktur von GBIF-Deutschland. Sie bildet eine der drei strukturellen Ebenen und hat vornehmlich informationstechnische Funktion. Es sollen Empfehlungen für die IT-Kommission erarbeitet, und dort entwickelte Fragestellungen bearbeitet werden. Die Aufgabe der IT-Fachgruppe ist das Sammeln und die Bereitstellung technischer Information zu IT-Entwicklungen im Bereich Biodiversitätsinformatik und die Veranstaltung mehrerer Workshops für Softwareentwickler. Es werden dabei nicht nur Projekte und Entwickler innerhalb der GBIF-Struktur angesprochen, sondern auch innerhalb anderer Organisationen bzw. Institutionen (z. B. BIOLOG). Ergebnisse des Workshops werden neben den Teilnehmern auch den Mitgliedern der IT-Kommission (GBIF-Deutschland) zur Verfügung gestellt und dienen als Hintergrundinformation und Dokumentationsmaterial für

Anträge. Ein wesentliches Ziel der Workshops besteht in der Unterstützung von Informationsaustausch zwischen den Teilnehmern. Hierbei sollen vor allem technische Aspekte im Vordergrund stehen, Evaluationen bzw. Stärken-Schwächen-Analyse durchgeführt, sowie Lösungsansätze und Änderungsmöglichkeiten herausgearbeitet werden.

### **Beiträge**

28.10.2004

- 12.45 Uhr „Datenaustauschformate für Sammlungsdaten: Das ABCD Schema“ (A. Güntsch, BGBM, Berlin)
- 13.30 Uhr „Daten und Datenflüsse in GBIF-D Botanik“ (A. Hahn, BGBM, Berlin)
- 14.15 Uhr „XML Sicherheitsmechanismen für den Datentransfer in GBIF-D“ (E. Langer und L. Suhrbier, FU Berlin)
- 15.00 Uhr „Webservices und XSLT“ (K. Schild, FU Berlin)
- 16.45 Uhr „Datentransfer als Scharnier zwischen ‚data mining‘ und ‚data consolidation‘ (G. Muche, Universität Hamburg)
- 17.15 Uhr „Datenflüsse in SysTax“ (Th. Ludwig, Universität Ulm)

29.10.2004

- 9.00 Uhr „Bericht über das TDWG Meeting 11.–17.10.2004, Christchurch, New Zealand“ (G. Hagedorn, BBA Berlin)
- 9.30 Uhr „Einsatz von Offline- und Online-Tools bei der Standardisierung und Qualitätssicherung der Erfassung von Namenslisten“ (A. Kroupa, Science4you, Bayreuth)
- 10.00 Uhr „Konzept zum Middleware-Framework innerhalb der *Diversity Workbench* und Einführung in XQuery, XPath und XPointer“ (C. Oertel, Universität Tübingen)

Die Präsentationen sind als pdf-Dateien auf der Homepage der IT-Fachgruppe hinterlegt ([http://132.180.63.25/GBIF-ITFG/WS\\_Beitraege\\_all.html](http://132.180.63.25/GBIF-ITFG/WS_Beitraege_all.html)).

### **Kurzfassungen der Beiträge**

#### **Datenaustauschformate für Sammlungsdaten: Das ABCD Schema**

Anton Güntsch

Das ABCD Schema (Access to Biological Collection Data) ist ein XML Schema für den Austausch biologischer Sammlungsdaten. Es wird seit dem Jahr 2000 durch

TDWG und eine CODATA Taskgroup (<http://www.bgbm.org/tdwg/codata/>) entwickelt und wird insbesondere durch BioCASE (<http://www.biocase.org>) und GBIF (<http://www.gbif.org>) gefördert.

Mit hunderten von Datenelementen deckt es Lebenssammlungen, konservierte Sammlungen und Beobachtungen aller taxonomischen Gruppen ab. Um Haltern von Sammlungsdatenbanken die Konfigurierung ihrer ABCD Netzwerksoftware zu erleichtern wurden Datenbereiche, von denen man annimmt, dass sie häufig in Datenbanken nicht ausreichend atomisiert wurden, sowohl atomisiert als auch als Freitext repräsentiert (variable Atomisierung). Zusätzlich wurde weitgehend auf den Einsatz kontrollierter Vokabulare und regulärer Ausdrücke verzichtet.

Da ABCD ausdrücklich nur Sammlungsdaten abdeckt, integrieren Biodiversitäts-Netzwerke ABCD Sammlungsdienste mit anderen Diensten. Der BioCASE / GBIF-D Prototyp nutzt zum Beispiel einen Registrierungsservice mit einer Indexdatenbank um ABCD Datenanbieter für eine konkrete Anfrage zu identifizieren. Synonyme werden mit einem taxonomischen Thesaurus mit in die Suche einbezogen. Verweisen die zurück gelieferten ABCD Dokumente auf Bilddaten, so werden diese mit einem speziellen Bild-Server dargestellt. Darüber hinaus können geographische Koordinaten von Fundorten über einen Service zur Erstellung von Verbreitungsdaten visualisiert werden.

### **Daten und Datenflüsse in GBIF-D Botanik**

Andrea Hahn

Der botanische Knoten von GBIF-D besteht aus zehn Teilprojekten, die neben der Datendigitalisierung selbst die Anbindung existierender und zu schaffender Datenbanken an ein gemeinsames Zugangssystem zum Ziel haben. Der gegenwärtige Schwerpunkt geförderter Projekte liegt auf Referenz- und Typussammlungen vorwiegend in Herbarien, zu denen ein Zugang in Form des „Deutschen Virtuellen Herbariums“ geschaffen wird. In jedem Fall sollen jedoch auch die anderen Disziplinen im botanischen Bereich (Phytodiversitätsforschung, Taxonomie, pflanzengenetische Ressourcen, Botanische Gärten) angemessen berücksichtigt und über ein gemeinsames Portal zugänglich gemacht werden. Dies bedeutet eine Zusammenführung von Beobachtungs- und Objektdaten sowie von taxonomischer (konzeptbasierter) Information, von Bild- und Textdaten, sowie die Schaffung von Kommunikationsstrukturen sowohl zwischen den beteiligten Projekten und Datenhaltern als auch zwischen Datenhaltern und Nutzern des Systems im Internet (Feedback).

Der Grundansatz folgt dabei dem von GBIF International propagierten Prinzip der dezentralen Datenhaltung: Datenbanken verbleiben bei ihren Erzeugern und werden über ein allgemeines Zugangssystem über das Internet abfragbar gemacht. Da sowohl die zugrunde liegende Technik als auch Datenbanksoftware und -strukturen typischerweise sehr heterogen sind, wird dabei der BioCASE-Ansatz angewendet,

Datenbanken über die Installation eines sogenannten Wrappers nach außen zu „konformieren“. Für die Datenübertragung sowohl der Nutzeranfragen als auch der Datenbank-Antworten werden das BioCASE-Protokoll und das ABCD-Schema (Access to Biological Collection Data) benutzt, sowie die vom BioCASE-Projekt dafür entwickelte Software. Protokoll und Schema werden auch von GBIF International unterstützt. Neben den angesprochenen Sammlungs- und Beobachtungsdaten wird eine taxonomische, konzeptorientierte Datenbank („Liste der Farn- und Blütenpflanzen“, Bundesamt für Naturschutz) dazu genutzt, Suchanfragen nach Namen automatisch auch auf verwandte Konzepte (Synonyme) auszuweiten. Die Registrierung angeschlossener Provider wird künftig auf die zentrale Providerdatenbank von BioCASE zugreifen und auch den dort gepflegten Thesaurus mit benutzen.

Ein eigenes Arbeitspaket in GBIF-D-Botanik befasst sich mit der Spezifikation und Implementierung eines sog. Fernannotations-Mechanismus: Die zunehmende Veröffentlichung von Bilddaten im Virtuellen Herbarium führt zu sich verändernden Arbeitsabläufen bei Kuratoren. Während im „physischen“ Herbar die Ergebnisse einer wissenschaftlichen Bearbeitung direkt am Objekt vermerkt werden (die sog. „Annotation“, z. B. bei einer Umbestimmung), ist im digitalen Herbar Feedback noch zumeist auf das Versenden von Emails oder die Eingabe in einfache Webformulare die Praxis. Die Besorgnis von Sammlungshaltern besteht vor allem darin, dass physische und virtuelle Sammlung zunehmend voneinander abweichende Informationen tragen, oder dass die Synchronisierung zu erheblicher Mehrarbeit auf kuratorischer Seite führt. Die künftige Gestaltung eines geeigneten Fernannotations-Mechanismus sowohl von technischer Seite als auch in Hinblick auf die Arbeitsabläufe in der Institution war Gegenstand eines Workshops im Februar 2004, in dem ein Testbetrieb für eine Erprobungsphase spezifiziert wurde. Inwieweit künftig auch ein Registrierungsverfahren für Annotations-Autoren eingeführt werden muss, ist zum gegenwärtigen Zeitpunkt noch offen. Das gleiche gilt bezüglich einer Entscheidung zur Speicherung von Annotationsdaten, zu Zugangsmodalitäten und ggf. einer Daten-Synchronisierung.

## **XML-Sicherheitsmechanismen für den Datentransfer in GBIF**

Ekaterina Langer

Unter dem Begriff Sicherheit wird im IT-Kontext der Schutz informationstechnischer Systeme vor beabsichtigten und mit Intelligenz betriebenen Angriffen verstanden. Sicherheit ist ein wichtiger Aspekt aller verteilten IT-Systeme, deren Nachrichten- und Datenkommunikation per Internet abgewickelt wird. Die übertragenen Daten und Nachrichten sind verschiedenen Bedrohungen ausgesetzt. Geheime Informationen können abgehört werden, Daten und Abfragen können manipuliert werden und Absenderdaten gefälscht, das Abschicken von Nachrichten und deren Inhalte kann abgestritten werden.

Der Vortrag gibt einen kurzen Überblick der Kernanforderungen der IT-Sicherheit – Vertraulichkeit, Integrität, Verbindlichkeit und Authentizität der Datenquellen- und der Techniken mit denen sie gewährleistet sind.

Das Projekt „XML-Sicherheitsdienste für GBIF-D“ hat zum Ziel die Entwicklung und Integration von Sicherheitsdiensten auf XML-Basis für das deutsche Biodiversitätsnetzwerk. Das Vorhaben befasst sich einerseits mit der Absicherung der Daten- und Nachfragekommunikation in GBIF-D, andererseits erzielt es die Erweiterung des Systems um Rechtemanagement und Zugriffskontrolle.

Um die Vorteile von XML-basierten Sicherheitsmechanismen im GBIF-Kontext zu verdeutlichen, werden die Besonderheiten der GBIF-Initiative und die technischen Eigenschaften der zugrunde liegenden BioCASE-Software diskutiert.

Folgende relevante XML-Sicherheitsstandards werden vorgestellt und ihre Anwendung anhand von Beispielen verdeutlicht:

XML Signatur

XML Encryption

XML Key Management Specification

eXtensible Access Control Markup Language

Security Assertion Markup Language

eXtensible rights Markup Language.

Anschließend wird der Einsatz der XML-basierten Sicherheitsmechanismen im Projekt anhand der erweiterten BioCASE-Architektur gezeigt.

## **Webservices und XSLT**

Klaus Schild

## **Datentransfer als Scharnier zwischen data mining und data consolidation**

Gerhard Muche

Das BIOTA AFRICA Projekt mit Head office am Biozentrum Klein Flottbek betreibt in Afrika eine umfangreiche Datenerhebung zur Biodiversität. Beteiligt sind über einhundert Wissenschaftler aus verschiedenen Disziplinen von unterschiedlichen Standorten und Ländern. Die Beobachtungen werden in der Regel elektronisch gespeichert, jedoch nicht in einem durchgehend einheitlichen Datenformat, sondern je nach Disziplin, persönlichen Gewohnheiten und verfügbaren Ressourcen mit etlichen handelsüblichen oder eigenentwickelten Softwareprodukten. Zweck der Datentransfer-Tools ist, die unterschiedlich angelieferten oder historisch gewachsenen Datenbestände in einem konsistenten Datenpool zu vereinigen. Ähnliche Aufgabenstel-

lungen finden sich in den vier GBIF-Projekten des Biozentrums Klein Flottbek der Universität Hamburg.

Das angewendete und bewährte Verfahren, die Datenbestände zusammenzutragen und zu homogenisieren, besteht darin, die Originaldaten via Import-/Exportfunktionen in ASCII- oder MS Access-Format zu transformieren und danach mit eigenentwickelten, kompilierten Programmen Datensatz für Datensatz zu interpretieren und in ein gewünschtes Zielformat zu bringen. Die Zielsysteme dienen einerseits als Instrumente zur weiteren Datenbereinigung und -aufbereitung durch Fachwissenschaftler und andererseits als Instrumente für die Internetveröffentlichung. Für die Sichtbarkeit im Netz steht dem Biozentrum Klein Flottbek das Rechenzentrum der Universität Hamburg zur Verfügung, auf dem MySQL und PHP bereitgestellt ist.

Interesse besteht an einem Erfahrungsaustausch, wie bestehende Datenbanken im Internet genutzt werden können, um weitgehend automatisch Aufgaben zur Datenhygienisierung ausführen zu können, z. B. um die Schreibweise der wissenschaftlichen Artnamen prüfen und ggf. korrigieren zu können.

### **Datenflüsse in SysTax**

Thorsten Ludwig

Dateneingabe:

- über die Client-Application (Forms6, Net8-Protokoll).
- Importe: flache Schnittstellendateien für mehrere Bereiche (Taxonomie, Synonymie, Sammlungsdaten, Multimedia)

Datenausgabe:

- über die Client-Application (Forms6, Net8-Protokoll).
- Exporte im Schnittstellenformat
- Web-Abfragen
- fremde Portale (z. B. GBIF-Wrapper)

### **Bericht über das TDWG Meeting 11.–17.10.2004, Christchurch, New Zealand**

Gregor Hagedorn

### **Einsatz von Offline- und Online-Tools bei der Standardisierung und Qualitätssicherung der Erfassung von Namenslisten**

Alexander Kroupa

## Konzept zum Middleware-Framework innerhalb der Diversity Workbench und Einführung in XQuery, XPath und XPointer

Clemens Oertel

Häufige Mehrfach-DNA-Sequenzierungen und fehlende Werkzeuge zur Speicherung der Ergebnisse von Sequenzanalysen motivierten die Erstellung einer eigenen Datenbank-Anwendung für molekularbiologische Daten ähnlich GenBank, jedoch mit diversen Erweiterungen. Hiervon unabhängig bestand Bedarf an einem System zur Veröffentlichung von Bildern und Zusatzinformationen im Internet (Handzeichnungen und TEM-Aufnahmen) sowie an einem zentralen Literaturverzeichnis. Das Ziel war somit ein Datenbanksystem zur Speicherung und Verknüpfung unterschiedlicher biologischer Daten. Operationen auf den gespeicherten Daten sollten der Erweiterbarkeit halber modular zu implementieren sein. Die Dateneingabe sollte per Browser geschehen, die -ausgabe sollte in verschiedenen Formaten (XML, PDF, GenBank, ...) möglich sein. Darüber hinaus sollte die Möglichkeit zur Einbindung bestehender legacy-Systeme bestehen. Im folgenden Entwicklungsprozess entstanden verschiedene Versionen einer entsprechenden Software: Einer in PHP und MySQL entwickelter Anwendung folgte eine C++-basierte Umsetzung, die weiterhin existierenden Schwächen führten schließlich zu einer in Eiffel implementierten Version mit PostgreSQL als Datenbanksystem. Die bis heute in Verwendung und Entwicklung befindliche Version basiert auf der üblichen modularisierten Trennung von Eingabe-, Ausgabe- und dazwischen befindlicher Verarbeitungsschicht. Die flexible Gestaltung der Dateneinbindung erlaubt Zugriffe auf in unterschiedlichsten Formaten vorliegenden Daten sowie auf externe Datenbanken, für die auch ein Caching-Mechanismus vorgesehen ist. Des Weiteren weist das System eine umfangreiche Nutzerverwaltung und Autorisierungsmechanismen auf. Der Workflow innerhalb des Systems für einzelne Anfragen kann flexibel festgelegt werden; es werden *hooks* bereitgestellt, an die verschiedenen Bearbeitungsschritte angehängt werden können. Die verwendeten Datenstrukturen entstammen dem *Diversity Workbench*-Projekt. Dessen Datentyp-Beschreibungen im XML Schema Format werden direkt zur Generierung der verwendeten Klassen eingesetzt. Darüber hinausgehende Datenstrukturen können durch eine Meta-Beschreibung zur Laufzeit dem System hinzugefügt werden (dies ist allerdings mit nicht unerheblichen Performance-Einbußen verbunden). Die Kommunikation zu anderen Modulen der *Diversity Workbench* findet über SOAP statt.

Aktuell findet eine Vereinheitlichung aller internen Pfadangaben auf XPath statt. Des Weiteren wird kontinuierlich an der Programmierung weiterer Daten-Input Module sowie einer weiteren Dynamisierung der Workflow-Festlegung gearbeitet.

Ein langfristiges Ziel ist der Ausbau zu einem Middleware-System, so dass eine komponentenweise Nutzung einzelner Funktionen möglich wird.



## ***Ergebnisse***

Es sollen konkrete Datenflussszenarien entstehen, in Form eines Flussdiagramms mit erklärendem Text, die als Dokumentation für Projektskizzen für Anträge an das BMBF verwendet werden können. Hier sollten vor allem andere Datentypen als Sammlungsdaten und Taxonomie im Vordergrund stehen. Für die Entwicklung der use cases kann das von Donald Hobern (GBIF international) entworfene use case-template verwendet werden (<http://www.gbif.org/Stories/STORY1040140792>):

***Use Case: <Descriptive name – includes an active verb>***

### **Actors**

<List of actors involved in use case – may  
be users, external systems or events>

### **Description**

<Goal to be achieved by use case>

### ***Flow of Events***

### **Pre-conditions**

<Any conditions that apply to execute the use case successfully>

### **Basic Flow**

<Sequence of interactions between actors and system required to achieve goal>

### **Post-conditions**

<Any conditions that apply after the use case has been executed>

### **Alternative Flows**

<Any alternative flows of events which may occur, including failure situations>

### ***Further Details***

### **Non-functional Requirements**

<Any non-functional requirements that the use case must meet>

### **Issues**

<Any issues that remain to be resolved>

History

Person	Date	Comments
<Original author>	<Creation date>	Created

***Vorschläge zu weiteren Workshops und Bereitstellung der Daten und Ergebnisse***

Der Workshop wurde sehr positiv bewertet, allerdings sollte bei den folgenden Workshops darauf geachtet werden, die Themen und Terminfestlegung so früh wie möglich und in einem größtmöglichen Kreis zu versenden.

Bei der Beantragung der Koordinationsstelle für die IT-Fachgruppe wurden folgende Themen für weitere Workshops vorgeschlagen: Web-Services (Wrapper, RPCs), Datentransfer-Sicherheitskonzepte, spezifische Probleme der Interoperabilität verschiedener Systeme (u. a. GIS Anbindung), Datenbanken für verschiedene Inhalte (z. B. deskriptive Daten), GBIF Deutschland und GBIF International: Integration, Interoperabilität. Es wurde mehrfach die Anregung gegeben, Workshops zu spezielleren Themen zu veranstalten, wo man z. B. die Möglichkeit hat eine neue Technik zu erlernen. Als mögliche Themen wurden vorgeschlagen: Management von Bilddaten und Multimedia-Daten, Referenzmanager, Migration von CFM nach JSP. Die hier vorgeschlagenen Themen wurden mit der Bitte um weitere Vorschläge im Dezember 2004 an die Ansprechpartner innerhalb GBIF gesendet.

Die Informationen zu den Workshops sollen frei über die Homepage der IT-Fachgruppe verfügbar sein. Es wird ein Personenverzeichnis der IT-Fachgruppen-Mitglieder (d. h. aller Workshopteilnehmer) erstellt. Zudem wird eine kommentierte Linkliste zu den Vorträgen der Workshop Teilnehmer zusammengestellt und über die Homepage verfügbar gemacht.

Auf der Homepage der IT-Fachgruppe finden sich neben den Informationen zu den Workshops auch zwei Online-Formulare, die allgemeine Informationen zu Institutionen und technische Details zu IT-Entwicklungen im Bereich Biodiversitätsinformatik abfragen. Die Mitglieder der IT-Fachgruppe werden gebeten die Formulare für die Entwicklungen in ihrem Zuständigkeitsbereich ausfüllen und absenden, um für das Statusseminar im Dezember in Bonn eine möglichst ausführliche Dokumentation zu ermöglichen.

Über die Aktivitäten der IT-Fachgruppe und Ergebnisse der Workshops wird in der nächsten Sitzung der IT-Kommission berichtet werden.

## **Protokoll zum Workshop III der IT-Fachgruppe (GBIF-D)**

### **„Konzept für eine Rich Client Plattform zur Integration von Diversity Workbench-Komponenten“**

Dieser Workshop fand in zwei Teilen statt, vom 6. bis 7. Januar und vom 27. bis 28. Januar 2005.

Beginn:	6. Januar 2005 10.00 Uhr bzw. 27. Januar 2005 10.00 Uhr
Ende:	7. Januar 2005 13.00 Uhr bzw. 28. Januar 2005 17.00 Uhr
Ort:	Botanische Staatssammlung, Menzinger Straße 67, 80638 München
Leitung:	Gerhard Rambold
Protokoll:	A. Kohlbecker, D. Neubacher, M. Weiss
Teilnehmer:	Andreas Kohlbecker (BBA Berlin), Dieter Neubacher (München), Wiltrud Spiesberger (BSM München), Dagmar Triebel (BSM München), Thomas Vesely (München), Markus Weiss (BSM München)

#### **Ziele**

Für die Antragstellung zur nächsten Projektphase von GBIF-Deutschland wird von den an IT-Entwicklungen beteiligten Institutionen ein auf bessere Synergiewirkung gerichtetes Konzept benötigt. Da in beiden, an der BSM, BBA und UBT angesiedelten Projekten „GBIF-D-Mykologie“ und „BIOTA Southern Africa – Teilprojekt S04“ *Diversity Workbench*-Komponenten entwickelt und eingesetzt werden, deren Implementierung bislang jedoch unterschiedlich erfolgt, erscheint eine bessere Abstimmung der Entwicklungsarbeit als dringend erforderlich. Ziel dieses Workshops war es deshalb, zu evaluieren, wie im Rahmen der beiden BMBF-Projekte, eine bessere Abstimmung bzgl. verwendeter Programmiersprachen, Entwicklungsumgebungen sowie eine bessere Koordination bei der Entwicklung von Anwendungen bzw. Arbeitsteilung bei der Erzeugung von Programm-Code erreicht werden könnte. Zur Diskussion stand dabei v. a. die Frage, ob für die derzeit heterogene *Diversity Workbench* das J2EE Framework als Basis genutzt werden könnte und der im Rahmen von Biota Southern Africa entwickelte Java-Client *Diversity Navigator* als Plattform dienen könnte.

## **Tagesordnung**

### **TOP 1**

Auswahl, Ergänzung und Reihenfolge der TOPs

### **TOP 2**

- a) Eignung des *Diversity Navigator* (DN) als *Diversity Workbench* (DWB)-Plattform
- b) Hintergrund: *Diversity Workbench*-Komponenten (M. Weiss, D. Triebel)
- c) An der BSM bereits realisierte SQL-Server-Installationen (M. Weiss, W. Spiesberger)
- d) *DiversityProjects* und DWB-mdb-Versions-Depot an der UBT (W. Spiesberger)
- e) Konzept und Architektur des *Diversity Navigator* (D. Neubacher)
- f) An der BSM bereits realisierte *Diversity Navigator*-DWB-PostgreSQL Konfigurationen (inkl. Produktionsweg: Access-DB-Modell → PostgreSQL-DB) (D. Neubacher, W. Spiesberger)
- g) Aktuelle Entwicklungen an der BBA (ASPX/JSP); Möglichkeiten der Integration der Aktivitäten (A. Kohlbecker)
- h) Welche Implementierungssprache und welches Framework wollen wir für die zukünftige Entwicklung einsetzen?
- i) Teamwork: Soll eine gemeinsame Entwicklung von Klassen/Komponenten angestrebt werden, die von allen beteiligten Entwicklern genutzt und im Team entwickelt werden? (Vermeidung von Mehrfachentwicklung ein- und derselben Funktionalität). Möglichkeiten der zentralen Versionsverwaltung z. B. per CVS?
- j) Diskussion
  - Möglichkeiten der Implementierung von DWB-Komponenten als DN-Plugins
    - o bzgl. flexibler Einbindung
    - o bzgl. „echter“ Interoperabilität (→ Webservices u. a. Mechanismen)
    - o als unabhängig laufende Applets und/oder Servlets
  - Dokumentation von Client-Architektur und Programmcode [UML-Diagramme und CRC-Cards]
  - *Diversity Workbench*-VB-Komponenten: Implikationen bei Migration nach .Net
  - *DiversityCollection* v. 2: Implikationen bei Migration von C# nach Java

- *Diversity Navigator* v. 1: Implikationen bei Migration nach J# oder C#
- Schnittstellen zwischen den Systemen. Möglichkeiten bei .Net bzw. Java
- Mögliche Arbeitsteilung und Workflow (CVS, DWB-mdb-Versions-Depot) und erforderliche Konventionen

### TOP 3

- a) Verwendung von PostgreSQL-Datenbanken als DWB-Datenspeicher
- b) aktuelle PostgreSQL-Features: Gegenüberstellung der Möglichkeiten bei SQL-Server und PostgreSQL (M. Weiss, A. Kohlbecker)
- c) Produktionsweg: *DiversityProjects* und DWB-Versions-FTP-Depot (bzw. mdb-Versions-Depot)
  - Mechanismus für routinemäßigen Neuaufbau aktualisierter DWB-PostgreSQL-Datenbank-Versionen
  - Möglichkeiten für Zwischendokumentation gerade aktueller DWB-Datenmodelle
- d) Integration zusätzlicher Datenbanksysteme (→ SQL-Server) als Hacks?
- e) DN im Offline-Einsatz

### TOP 4

Revision von Szenario für Arbeitsteilung und Workflow (CVS, .mdb-Versions-Depot) und erforderliche Konventionen

### TOP 5

Strategien für Migration unter Berücksichtigung des Erhalts bereits bestehender Funktionalitäten.

## ***Ergebnisse***

### **1 Hintergrund**

#### **Einsatz von Diversity Workbench-Komponenten in den Projekten GBIF-D und BIOTA Southern Africa**

GBIF-D: a) MS Access-Applikationen mit den Komponenten: *DiversityDescriptions*, *DiversityCollection*, *DiversityReferences*, *DiversityGazetteer*, *DiversityExsiccatae*, *DiversityUsers*, *DiversityMapIndexing*; b) .Net-Applikationen (C#): *DiversityResources*, *DiversityUsers*, *DiversityCollection*, *DiversityTaxonNames*

BIOTA Southern Africa: c) Diversity Navigator Datenbank-Client (Java-Applikationen) mit den Komponenten: *DiversityDescriptions*, *DiversityCollection*, *DiversityChecklists*, *DiversityTaxonNames*

### **Aufwand für Portierung vorhandener Anwendungen bzw. Klassenbibliotheken von Diversity Workbench-Komponenten**

#### **C# → Java**

*DiversityCollection* v. 2.0 (Aufwand: hoch); *DiversityResources* (Aufwand: mittel) + *DiversityUsers* (Aufwand: minimal); evtl. Ersatz von *DiversityUsers* durch LDAP; *DiversityTaxonNames* (Aufwand: mittel). Entscheidung zwischen Eclipse und NetBeans. (Falls die Entscheidung endgültig zugunsten von Java fällt, wäre zu überlegen, welches Framework eingesetzt werden sollte: reines JSP (vs. J2EE), Velocity, Struts etc.).

#### **Java → C#**

*Diversity Navigator* (Aufwand: hoch).

## **2 Konzept**

Redesign und Implementierung der *Diversity Workbench* soll nach den Kriterien Plattformunabhängigkeit und Open Source erfolgen, welche als Voraussetzungen für optimale Kooperationsmöglichkeiten mit weiteren Arbeitsgruppen angesehen werden. Für eine Implementierung bieten sich grundsätzlich .Net und Java als Optionen an. Beide Alternativen wurden im Detail diskutiert. Dabei wurde festgestellt, dass der Aufwand bei Migration/Portierung für beide Richtungen etwa vergleichbar wäre. Seit Februar 2005 wird von der NSF ein Projekt (Direktor: Jim Beach, University of Kansas) zur Portierung der Datenbankanwendung *Specify* nach Java gefördert. Die Teilnehmer des ITFG Workshop 3 kamen nach Kenntnisnahme dieser Information überein, dass die daraus sich ergebenden Kooperationsmöglichkeiten genutzt werden sollten und deshalb einer Implementierung der *Diversity Workbench* in Java der Vorzug gegeben werden sollte.

### **Architektur**

Entsprechend bewährter Konzepte, wird auch für die anvisierte Implementierung einer *Diversity Workbench* eine dreischichtige Architektur als am besten geeignet angesehen.

Siehe auch Argumentation unter Punkt 3.4.1.7: „Sonstige Anmerkungen“ von A. Kohlbecker.

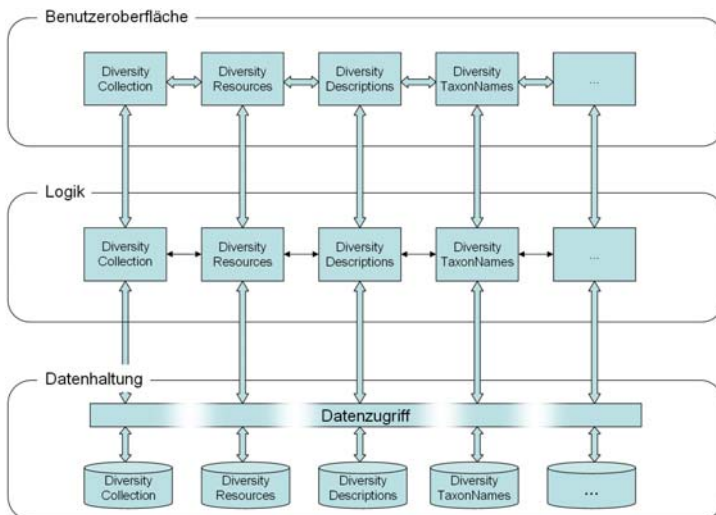


Abb. 1. Architektur des Workbench-Frameworks.

### Local Access

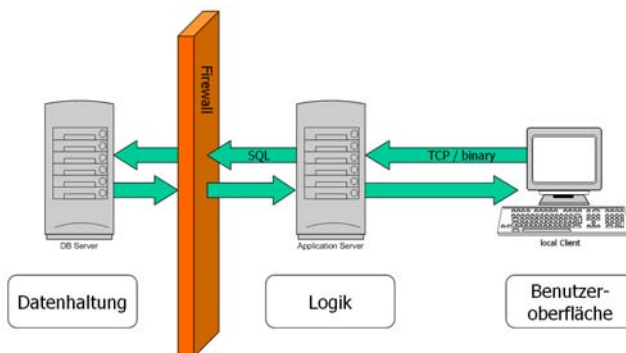


Abb. 2. Local Access.

### Remote Access

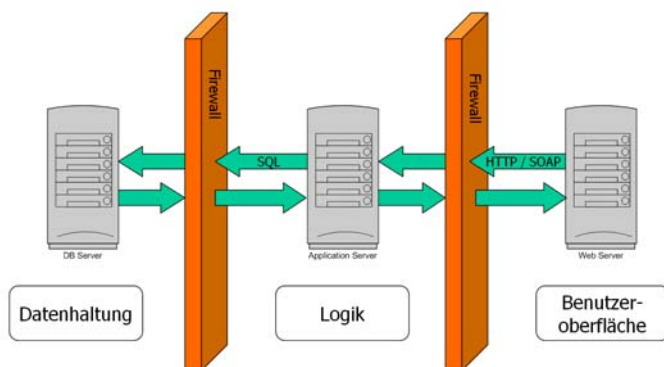


Abb. 3. Remote Access.

## Datenbanken

Es wird davon ausgegangen werden, dass von den verschiedenen Daten Providern auch künftig Datenbestände in unterschiedlichen Datenbanksystemen unter verschiedenen Betriebssystemen (und Formaten) bereitgestellt werden. Aus diesem Grunde erscheint es nicht sinnvoll, für die Datenhaltung ein breiteres Spektrum relationaler, für den Anwender transparenter (und auch als Stand-alone laufender) Datenbanksysteme, deren Strukturen den in der *Diversity Workbench* für die jeweiligen Komponenten publizierten Datenmodellen folgen, vorzusehen. Zunächst in besonderem Maße unterstützt werden sollen die Datenbanken PostgreSQL 8.0 (Open Source: Linux, Windows) sowie MS SQL-Server (Windows), mittelfristig ergänzt durch DB2 (Unix) und Oracle (Unix, Windows). Trigger bzw. Stored Procedures können je nach System als Java Stored Procedures (PostgreSQL, Oracle, DB2) oder in C# oder VB (SQL-Server, MS Access) programmiert sein. Aus Gründen der Sicherstellung von Betriebssystemunabhängigkeit und Datenbankunabhängigkeit sollte Hibernate als Persistenzmanager verwendet werden.

## Programmierung von Geschäftslogik-Layer und Client in Java

Geschäftslogik-Layer: Gewisse Teile der Business-Logik sollen unabhängig von Datenbank und Client als Mittelschicht implementiert werden. Zur Vermeidung von Redundanzen bei den Entwicklungsaktivitäten bei gleichzeitiger Betriebssystemunabhängigkeit soll als Programmiersprache ausschließlich Java verwendet werden. Die direkte Interoperabilität zwischen den verteilten Datenbank-Komponenten soll über JDBC-Schnittstellen erreicht werden. Als Webserver wird Apache mit Tomcat und JSP favorisiert. (Vorhandene IIS-Anwendungen sollten portiert werden.)

GUI: Ebenfalls aus Gründen der Sicherstellung von Betriebssystemunabhängigkeit sollen die *Diversity Workbench*-Komponenten-Clients (→ „*Diversity Navigator*“) als Plug-ins in Java realisiert werden, wobei als Rich Client Platform (RCP) a) Eclipse mit SWT als GUI dienen könnte; alternativ b) Eclipse mit Swing als GUI oder c) NetBeans mit Swing als GUI in Frage kommen könnten.

Für die Authentifizierung ist LDAP als separater Service vorgesehen.

## Schnittstellen

Schnittstellen zwischen den Modulen; Interfaces im Sinne von OOP. Würde man sich darauf einigen, alle zukünftigen Entwicklungen streng nach dem MVC (Model/View/Controller) Konzept zu entwickeln, bräuchten im Prinzip die M- und C-Komponenten nur einmal geschrieben werden, um dann für eine Standalone- oder Webanwendung lediglich die speziellen Views separat zu programmieren. Wäre es sinnvoll die *Diversity Workbench*-Komponenten als ThinClients aufzubauen? Zu bedenken ist hierbei, ob sich der Mehraufwand der durch eine MVC-Architektur entsteht



sich tatsächlich rechnet, d. h. ob letztlich die Synergieeffekte den Mehraufwand rechtfertigen.

Um eine Trennung zwischen der graphischen Repräsentation und der eigentlichen Funktionalität und Interaktion des GUI's zu erreichen, sollte das GUI nach dem Model View Control Pattern (MVC2) realisiert werden.

## Protokoll zum Workshop IV der IT-Fachgruppe (GBIF-D)

### „Integration und Anbindung von BIOTA-Africa Forschungsdaten in das GBIF-Netzwerk“

Beginn:	12. Mai 2005	12.00 Uhr
Ende:	13. Mai 2005	14.00 Uhr
Ort:	Botanische Staatssammlung, Menzinger Straße 67, 80638 München	
Leitung:	Gerhard Rambold	
Protokoll:	Alexandra Kehl	
Teilnehmer:	Anna Augustin (Herbarium Hamburgense, Universität Hamburg), Stefan Dressler (Forschungsinstitut Senckenberg, Frankfurt a. M.), Andrea Hahn (BGBM Berlin), Jörg Holetschek (BGBM Berlin), Jan Ingenhaag (BSM München), Manfred Keil (DLR Oberpfaffenhofen), Karl-Heinz Lampe (Museum Alexander Koenig, Bonn), Thorsten Ludwig (Universität Ulm), Gerhard Mucho (BIOTA AFRICA, Universität Hamburg), Marco Schmidt (Forschungsinstitut Senckenberg, Frankfurt a. M.), Christoph Schultz (Universität Würzburg), Melanie Vogel (Universität Würzburg), Markus Weiss (BSM München)	

#### **Ziele**

Auf diesem Workshop soll ein Überblick über Datenstrukturen und Datenflüsse in den verschiedenen Teilprojekten von BIOTA AFRICA erhalten werden und des Weiteren Datenflussszenarien entworfen werden, die für die nächste Antragsphase für GBIF-D genutzt werden können. Diese Datenflüsse oder „Use Cases“ sollen die Voraussetzungen und grundlegenden notwendigen Schritte für den Transfer unterschiedlicher Datentypen zeigen.

#### **Beiträge**

12.5.2005

12.30 Uhr „Für GBIF Deutschland verwendete Technologien: das ABCD Schema und die BioCASE Technologie“ (J. Holetschek, A. Hahn, BGBM Berlin)

- 13.15 Uhr „Anspruch der Biodiversitätsforschung an Daten als Argumentationsgrundlage“ (G. Muche, BIOTA AFRICA, Universität Hamburg)
- 13.45 Uhr „Die Anbindung des Herbarium Hamburgense an GBIF als Ausgangslösung einer online BIOTA Datenbasis“ (Anna Augustin, Herbarium Hamburgense, Universität Hamburg)
- 14.45 Uhr „Wrapper Installation am GBIF Knoten Mykologie“ (Jan Ingenhaag, Wiltrud Spiesberger, Markus Weiss, BSM München)
- 15.15 Uhr „SysTax“ (Jürgen Hoppe, Thorsten Ludwig, Universität Ulm)
- 15.45 Uhr „Datenbanken und Datenbankclient Diversity Navigator, BIOTA Süd“ (Alexandra Kehl, Luciana Zedda, Universität Bayreuth)
- 16.15 Uhr „VegDa – eine Datenbank für Vegetationsaufnahmen und botanische Inventare“ (Marco Schmidt, Stefan Dressler, Forschungsinstitut Senckenberg, Frankfurt a. M.)
- 17.15 Uhr „BIOTA E15a: Datenerfassung von BIOTA-Ostafrika Belegexemplaren und ihre Integration in BioCASE und GBIF“ (Karl-Heinz Lampe, Museum Alexander Koenig, Bonn)
- 17.45 Uhr „Datenstrukturen in den Biota Teilprojekten W01 und S01“ (Manfred Keil, DLR Oberpfaffenhofen, Christoph Schultz, Universität Würzburg)

Die Präsentationen sind als pdf-Dateien auf der Homepage der IT-Fachgruppe hinterlegt ([http://132.180.63.25/GBIF-ITFG/WS\\_Beitraege\\_all.html](http://132.180.63.25/GBIF-ITFG/WS_Beitraege_all.html)).

## **Ergebnisse**

### **1. Beispielhafte Zuordnung (Mapping) Datentypen aus Biota Base auf das bestehende ABCD-PyWrapper-Design**

Zu Beginn wurde ein Datenschema entwickelt, welches die Datenstruktur von BIOTA aufzeichnet, um sie mit dem bestehenden, auf das ABCD-Schema angepassten Wrapper zu vergleichen.

Die Eigenheiten beschreibender Daten wird diskutiert. Sie unterscheiden sich grundlegend von den reinen Bestandsdaten, hier hinein fallen auch sozioökonomische und teilweise ethnobotanische Daten. Auch pflanzensoziologische Untersuchungen passen nicht in das ABCD-Schema. Hier müssten andere Protokolle benutzt werden.

Gerhard Rambold äußert Bedenken, ob es zweckmäßig sei, beschreibenden Daten in derselben Datenbank zu verwalten wie Belegdaten, trotz eines engen Bezuges könnten diese Bereiche nebeneinander stehen. Andrea Hahn meint, dass man die Belegdatenbanken auch in Hinblick auf beschreibende Daten wie Deckung erweitern

könnte. Das ABCD-Schema ist objektbezogen und müsste erweitert werden. Karl-Heinz Lampe führt an, dass Zusatzinformationen zu Belegen sehr divers sind, gerade wenn in mündlicher Form als Zusatzinformation in der Datenbank angeführt. Christoph Schultz schlägt ein Modell als Lösung vor, welches Zusatzinformationen hierarchisch an Belegdaten mit einem Rasterbezug bindet. Gerhard Muche führt an, dass in BIOTA bereits verschiedene Disziplinen mit einer einheitlichen Nomenklatur verlinkt werden.

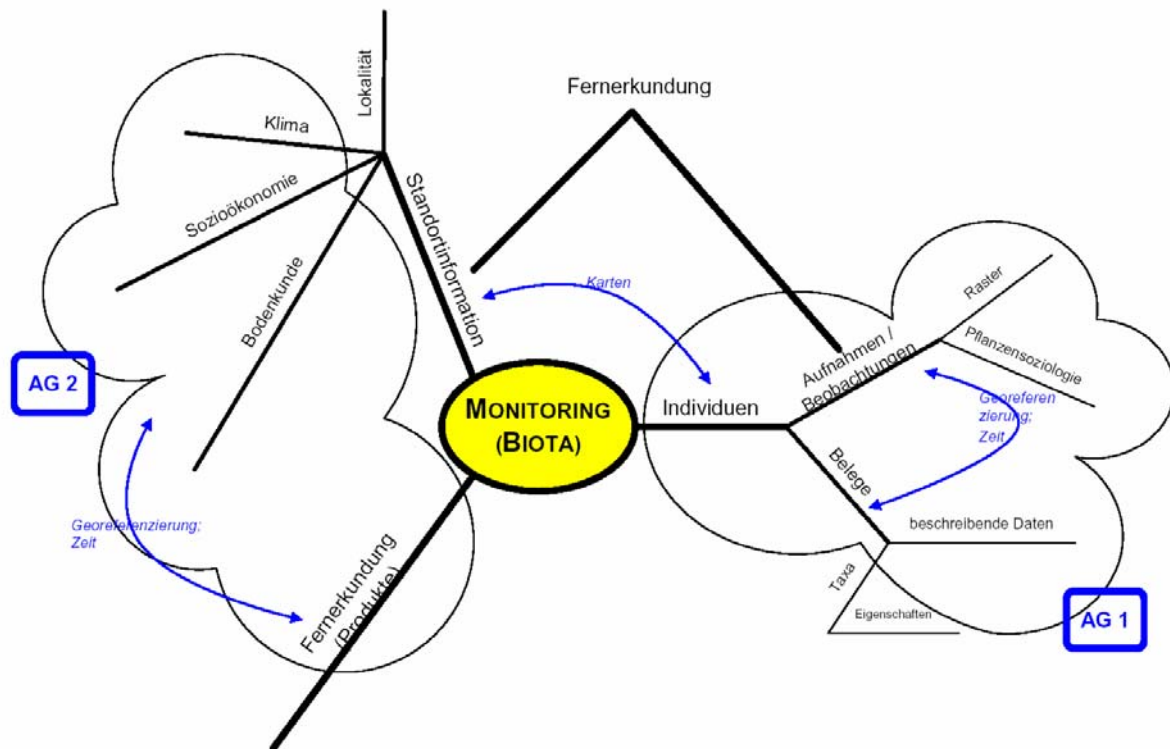


Abbildung 1: Mindmap zu Datenbereichen in BIOTA AFRICA (Andrea Hahn).

Durch das Benennen und die Analyse der Datenbereiche in BIOTA AFRICA konnten zwei Arbeitsgruppen (AG1 und AG2 vgl. Abbildung 1) identifiziert werden. Aus der Notwendigkeit, die hohe Diversität an Datenstrukturen, die in BIOTA Ost, West und Süd aufgenommen werden, zu verwalten und schließlich im Internet publizieren zu können, wurden im Rahmen dieses Workshops drei weitere „Miniworkshops“ mit folgenden Inhalten geplant:

1) Fernerkundungs – Standortdaten – abiotische Daten

Koordinator: Melanie Vogel und Christoph Schultz, Termin: 22.-23. Juni 2005, Universität Würzburg.

Teilnehmer (vorgesehen): Gerhard Muche, Markus Weiss, Tanja Krauß E02, Andreas Kohlbecker (angefragt), Alexandra Kehl (nur 2. Tag), Günter Miehllich oder Andreas Petersen S02 (angefragt).

Inhalt/Ziel: Überlegung wie man den Top-down-Ansatz der Fernerkundung am besten mit dem Bottom-Up-Ansatz der Standortkundler verschneiden kann. Ziel: Datenmodell für die in BIOTA erfassten Standorteigenschaften (u.a. Fernerkundung, Bodenkunde, Klimatologie).

2) Belegdaten – Vegetationsdaten

Koordinator: Gerhard Muche, Termin: 12. bis 15. Juli 2005, Forschungsinstitut Senckenberg, Frankfurt a. M..

Teilnehmer: Jörg Ewald (angefragt), Marco Schmidt, Markus Weiss, Andrea Hahn, Lothar Menner (angefragt).

Inhalt/Ziel: Ist ein gemeinsames Datenmodell für BIOTA Sammlungs- und Vegetationsdaten möglich?

3) Zusammenführung der Ergebnisse der ersten beiden Workshops

Koordinator: Karl-Heinz Lampe, Termin: 1.-2. September 2005, Museum Alexander Koenig, Bonn.

Als Teilnehmer sind die Teilnehmer der vorangegangenen Workshops vorgesehen sowie Vertreter der Sozioökonomie in BIOTA.

Zusammenführung der Ergebnisse, dabei wird ein beispielhaftes Mappen der angedachten Datenbanksysteme auf das CIDOC-CRM-Schema geschehen mit dem Ziel der Erstellung einer Präsentation für das BIOTA Statusseminar im November 2005.

Dabei sollen die Vertreter der Teilprojekte darauf achten, dass die Anforderungen der afrikanischen Counterparts berücksichtigt werden.

Zusätzlich soll ein Observatorium ausgewählt werden, an dem beispielhaft die Möglichkeiten der Verschneidung der abiotischen und biotischen Daten aus den unterschiedlichen Fachgebieten erprobt werden soll. Dazu soll Ute Schmiedel kontaktiert werden.

## **2. Identifikation noch fehlender Bereiche im ABCD-Schema für den BIOTA-internen Gebrauch**

Weiterhin wurden fehlende Bereiche im ABCD-Schema identifiziert. Es wurde folgende Tabelle 1 erstellt.

Tabelle 1: Datentypen aus BIOTA AFRICA und Abdeckung durch des ABCD-Schema.

<b>BIOTA AFRICA</b>	<b>ABCD</b>	<b>GBIF</b>
Standortdaten	Gathering event	
Vegetationsaufnahmen	Bestimmungen Aufsammlungsereignis	
Bodenkunde	Evtl. Biotope data	
Landnutzung	Keine Möglichkeit	
Pflanzensoziologie	Synökologie Biotopmeasurement	
Deskriptive Daten		
Fernerkundung		Metadaten => Verfügbarkeit Zugreifbarkeit

Eine Gegenüberstellung der BIOTA AFRICA Daten mit dem ABCD-Schema ergab dass außer den Bereichen „Belege“ und „Einzelbeobachtungen“ alle Bereiche in ABCD nur zusätzlich und zwanglos abgebildet werden können. Die Diskussion ergab, dass andere bereits existierende Transferstandards (z. B. EML: <http://knb.ecoinformatics.org/software/eml/>, GML: <http://www.opengis.net/gml/>) evaluiert werden müssen. Für die BIOTA AFRICA Forschungsdaten soll ein modulares System angestrebt werden, die Schnittstellendefinitionen sollen deshalb im Vordergrund stehen. Herr Lampe stellt kurz CIDOC-CRM vor, das als grundlegendes Konzept für Datenstrukturen genutzt werden kann und als (in Kürze) internationaler Standard die Definition der Schnittstellen zwischen den verschiedenen Modulen erleichtern kann.

## Protokoll zum Workshop V der IT-Fachgruppe (GBIF-D)

### „Erstellung eines Datenmodells für die in BIOTA erfassten Standorteigenschaften (Fernerkundung, Bodenkunde, Klimatologie)“

Beginn:	22. Juni 2005	13.00 Uhr
Ende:	23. Juni 2005	15.00 Uhr
Ort:	Universität Würzburg, Institut für Geographie, Am Hubland, 97074 Würzburg	
Leitung:	M. Vogel, G. Muche	
Protokoll:	M. Vogel	
Teilnehmer:	Minnattallah Boutros (Universität Würzburg), Alexander Gröngröft (Universität Hamburg), Andreas Heinecke (FH Gelsenkirchen), Alexandra Kehl (Bayreuth, ITFG-Koordination), Karl-Heinz Lampe (Museum Alexander Koenig, Bonn), Gerhard Muche (BIOTA AFRICA, Universität Hamburg), Christoph Schultz (Universität Würzburg), Melanie Vogel (Universität Würzburg), Markus Weiss (BSM München)	

#### **Ziele**

Im Rahmen dieses Workshops soll ein Datenschema für eine Datenbank erarbeitet/gesucht werden, in dem sich Standorteigenschaftsdaten der BIOTA-Teilprojekte S01/W01 (Fernerkundung), S02 (Bodenkunde) und W02 (Klima) vereinen lassen. Die GBIF IT-Fachgruppe trägt mit ihren Erfahrungen aus dem GBIF-D-Projekt Informationen zu Möglichkeiten der Datenhaltung bei.

#### **Beiträge**

22.6.2005

14.00 Uhr „Struktur der Bodenkunde-Daten“ (A. Gröngröft, Universität Hamburg)

15.00 Uhr „Strukturen Fernerkundungsdaten (u. Klimadaten)“ (M. Vogel, Universität Würzburg)

15.30 Uhr „Erfahrungen bei der BIOTA-Datenhaltung“ (G. Muche, BIOTA AFRICA Universität Hamburg)

16.00 Uhr „Weitere Möglichkeiten der Datenhaltung“ (C. Schultz, Universität Würzburg)

17.00 Uhr „Datenbanken der *Diversity Workbench*“ (M. Weiss, BSM München)

Die Präsentationen sind als pdf-Dateien auf der Homepage der IT-Fachgruppe hinterlegt ([http://132.180.63.25/GBIF-ITFG/WS\\_Beitraege\\_all.html](http://132.180.63.25/GBIF-ITFG/WS_Beitraege_all.html)).

### **Verlauf der Diskussion und Ergebnisse**

Als Kernfragen wurden identifiziert:

- Wer soll der Anwender der resultierenden Datenbank (DB) sein?
- Sollen die Standortdaten in dieser DB Realdaten oder Metadaten sein?
- Sollen ggf. Realdaten zur reinen Visualisierung von Fundorten dienen oder zur GIS-Analyse?
- Soll die DB der langfristigen Datensicherung und -bereitstellung dienen?
- Was wäre unter Berücksichtigung aller Punkte die Optimal- und was die Minimallösung?
- Welchen Wert hätte bei diesen Überlegungen das CIDOC CRM-Modell (<http://cidoc.ics.forth.gr/>)?
- Welche Schnittstellen müssen zu anderen Datenbanken, GBIF etc. vorhanden sein?
- Des Weiteren wurde nochmals der Mehrwert einer Zusammenarbeit von BIOTA und GBIF betont, durch die eine nachhaltige Datenvorhaltung und die Basis zur Analyse der biologischen BIOTA-Funddaten und auch GBIF-Belegdaten ermöglicht werden wird.

### **Struktur der Bodenkunde-Daten (A. Gröngröft)**

Das Bodenkunde-Projekt hält eigene Analysedaten vor, sowie weitere Fremddaten als Raster- oder Vektor-Dateien, wie z. B. die FAO-Bodenkarte.

### **Struktur der Fernerkundungsdaten (u. Klimadaten) (M. Vogel)**

Es besteht für die Fernerkundungs- und GIS-Daten bereits eine Metadatenbank, die für die hier diskutierte Fragestellung weiter ausgebaut werden kann. Die Kernfrage wäre auch hier, ob die Datenbank dann „verfügbare Daten“ (also Metadaten am Abfragepunkt) oder „verfügbare Werte“ (also Dateninformation am Abfragepunkt) beithalten soll.



### **Weitere Möglichkeiten der Datenhaltung (C. Schultz)**

Es werden verschiedene Möglichkeiten zur Datenhaltung vorgestellt. Es werden GIS-Systeme der ersten (fest zugeordnetes System), zweites (loses Verbundsystem) und dritten Generation (integriertes System) vorgestellt. Bedingung für den Aufbau und den interdisziplinären Betrieb von Systemen der zweiten und dritten Ordnung ist allerdings eine Berücksichtigung internationaler für die jeweilige Domäne vorliegenden Metadaten-Standards. Für die Domäne der Fernerkundung werden hier die exemplarisch die Modelle EML – Ecological Metadata Language in Verbindung mit SML – Sensor Model Language vorgestellt.

In der anschließenden Diskussion wird auch über eine Integration der räumlichen Daten in existierende Datenhaltungssystem DIMS des Deutschen Fernerkundungsdatenzentrum (DFD) gesprochen. Aufgrund der knappen Ressourcen soll daher auch eine mögliche Integration der räumlichen BIOTA-Daten in dieses bestehende System geprüft werden.

([http://www.caf.dlr.de/caf/technologie/technologie\\_transfer/technologie\\_transfer\\_caf/technologie\\_transfer\\_dims/\\_technologie\\_transfer\\_dims](http://www.caf.dlr.de/caf/technologie/technologie_transfer/technologie_transfer_caf/technologie_transfer_dims/_technologie_transfer_dims))

### **Erfahrungen bei der BIOTA-Datenhaltung (G. Muche)**

Die Diskussion rankt sich um folgende Fragen:

Sollten derzeit unbenutzte Datenfelder aus der DB eliminiert werden, um den Datenbestand „übersichtlich“ zu halten oder sollten sie für zukünftige Ansätze weiterhin bereitstehen.

Sollten zusätzliche Datenfelder, z. B. aus dem Metadatensatz der Satellitenszenen in die DB implementiert werden, auch wenn sie derzeit irrelevant erscheinen. Hintergrund ist die Vermeidung des Nachpflegens von Datensätzen, wenn später Bedarf an diesen Informationen entstehen sollte.

Konsens besteht darin, dass offenbar die Wahrnehmung von Information und die Bewertung der Relevanz der Daten in den einzelnen Teilprojekten extrem verschieden sind. Es herrscht bei den Teilnehmern die Grundstimmung, soviel Information wie möglich in die DB eingehen zu lassen.

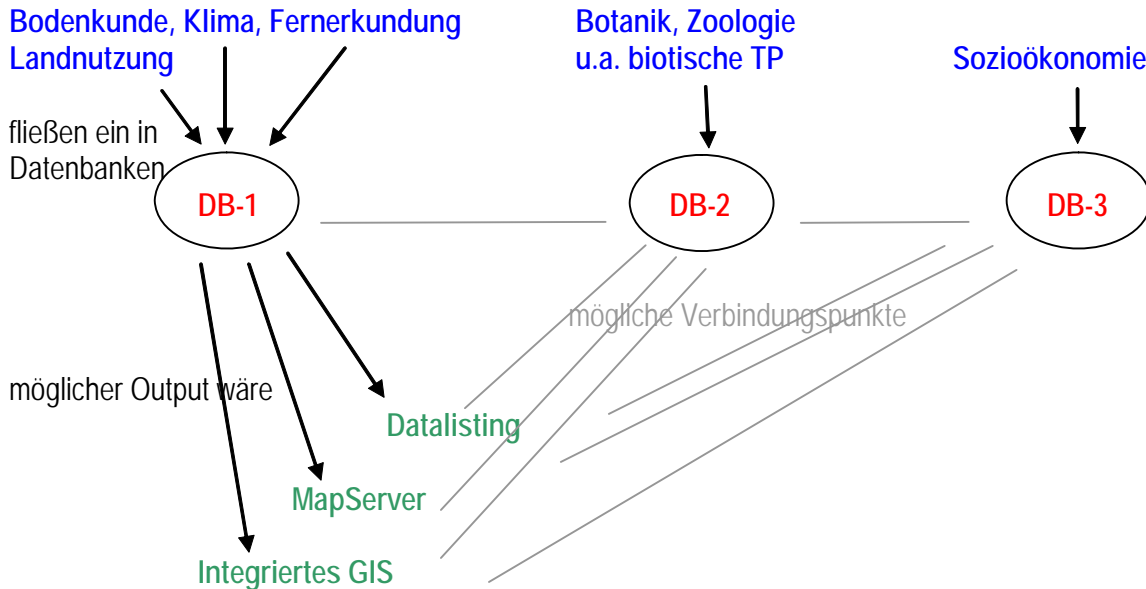
### **Datenbanken der *Diversity Workbench* (A. Weiss)**

Die Modellierung des Waldwachstums in Bayern (SILVA) und das ABCD (Access on Biological Collection Data)-Schema werden vorgestellt.

## Fortführung Diskussion und Konkretisierung des zu entwickelnden Daten-schemas

Die verschiedenen möglichen Ziele einer BIOTA-Datenbankstruktur werden grob von M. Vogel skizziert:

Information aus



Dabei liefert das Data-Listing als Output die Information, welche Daten zu einem Abfrageort vorhanden sind (Metadaten). Mit dem MapServer lassen sich Abfragepunkte auf vorhandenen GIS-Datensätzen, wie z. B. Vegetationskarten, Höhenmodellen oder NDVI-Daten visualisieren. Das integrierte GIS erlaubt die Analyse der Abfragepunkte anhand von GIS-Daten.

Es kann zunächst jedoch keine Einigung erzielt werden, welche der drei Möglichkeiten – oder welche Kombination daraus – angestrebt werden sollte.

## CIDOC CRM

K.-H. Lampe stellt das konzeptionelle Datenmodell CIDOC CRM vor. Dabei handelt es sich um ein Conceptual Reference Model (CRM) des International Committee for Documentation (CIDOC) (→ [http://cidoc.ics.forth.gr/docs/cidoc\\_crm\\_version\\_4.0.pdf](http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.0.pdf)).

Dieses Modell setzt alle Daten in Abhängigkeit von „Akteur“, „Ereignis“ und „Zeit“. Dieses Schema mit seiner sog. „Event-Centric“ soll zur ISO-Norm erhoben werden. Seine Anwendung stellt damit keine Insellösung dar, sondern einen übertragbaren Standard, der die potenzielle Verknüpfung mit fremden Datenbeständen ermöglicht.

Es wird die Übertragbarkeit des Modells für BIOTA-Zwecke diskutiert. In der Präsentation liegt der Fokus auf „Akteur“ und „Zeit“, jedoch können auch Sammelobjekte dargestellt werden.

Satellitenbilder und Metadaten können analog dem Beispiel für Fotos eingegeben werden. Die Bearbeitung von Satellitendaten (Georektifikation etc.) wäre demnach ein Zerstörungs- oder Modifikations-Event. Profilgrabungen der Bodenkundler wären ebenfalls Events usw.

Fragestellungen wie die mögliche Umsetzung eines Web-Access bleiben von CIDOC CRM völlig unberührt, denn CIDOC CRM ist dabei keine „Vorlage“ oder „Blanko-Datenbank“, sondern lediglich ein Schema, in dem man zur Überprüfung von Datenbeständen die Daten zu einander in Beziehung setzen kann. Die sehr heterogenen Datenbestände der BIOTA-Teilprojekte könnten durch die CIDOC-Analyse effizient homogenisiert werden. Die Daten erhalten durch dieses Schema eine Struktur, die die Nutzung der Daten auch Außenstehenden (nach Projektende) ermöglicht, die mit den bisherigen BIOTA-Strukturen der einzelnen Teilprojekte nicht vertraut sind.

Zudem werden die Daten der Teilprojekte auch für die anderen Projektpartner transparent.

Konsens ist, dass die Datenstrukturierung in BIOTA/GBIF konzeptionell mit CIDOC CRM weiter verfolgt werden sollte.

### **Ausblick auf den 7. Workshop der IT-Fachgruppe in Bonn; Aufgabenverteilung**

Zur Vereinfachung der Sprechweise wird für das zu erstellende Datenmodell in BIOTA ein Name gesucht. Vorläufiger Konsens für das entstehende räumliche Content Manager System (rCMS) ist die Bezeichnung

„BIOTA-Knowledge-Organisation System (B-KOS oder BIOTA-KOS) in Terms of CIDOC CRM“

Aufgaben für die Teilnehmer und BIOTA-Teilprojekte (TP):

- G. Muche trifft sich mit K.-H. Lampe und/oder A. Heinecke um die Möglichkeiten von CIDOC anhand von konkreten BIOTA-Datenbeispielen weiter auszuloten.
- G. Muche stellt die Ergebnisse im nächsten Workshop am 14.-15. Juli in Frankfurt den biotischen Teilprojekten vor.
- G. Muche entwirft außerdem ein Positionspapier, in dem anhand der Beispieldaten die Vor- und Nachteile von CIDOC für BIOTA dargestellt werden.

## Protokoll zum Workshop VI der IT-Fachgruppe (GBIF-D)

### „Datenbanken zur Verwaltung ökologischer Daten in BIOTA und GBIF am Beispiel botanischer und zoologischer Beobachtungsdaten: Anforderungen, Gemeinsamkeiten, Perspektiven“

Beginn:	14. Juli 2005	13.00 Uhr
Ende:	15. Juli 2005	15.00 Uhr
Ort:	Forschungsinstitut Senckenberg, Frankfurt a. M., Senckenberganlage 25, 60325 Frankfurt	
Leitung:	Gerhard Mucbe	
Protokoll:	Alexandra Kehl	
Teilnehmer:	Minnattallah Boutros (Universität Würzburg), Jörg Ewald (FH Weihenstephan), Manfred Finckh (Universität Hamburg), Frauke Fischer (Universität Würzburg), Andrea Hahn (BGBM Berlin), Andreas Heinicke (FH Gelsenkirchen), Alexandra Kehl (Bayreuth, ITFG-Koordination), Karl-Heinz Lampe (Museum Alexander Koenig, Bonn), Rudolf May (BfN, Bonn), Lothar Menner (Forschungsinstitut Senckenberg, Frankfurt a. M.), Gerhard Mucbe (BIOTA AFRICA, Universität Hamburg), Johannes Penner (Universität Würzburg), Marco Schmidt (Forschungsinstitut Senckenberg, Frankfurt), Peter Strobl (Salzburg), Markus Weiss (BSM München)	

#### **Beiträge**

14.7.2005

13.30 Uhr „Überblick zur Diskussion der laufenden Workshop-Serie (ITFG-WS IV–VII)“ (G. Mucbe, Hamburg)

14.00 Uhr „Struktur der DWB-Datenbankkomponente DiversityEcology“ (M. Weiss, München)

14.45 Uhr „Vegetationskundliche Daten des Impetus-Projektes“ (M. Finckh, Universität Hamburg)

15.15 Uhr „Datenvielfalt in der Vegetationsdatenbank VegDa (BIOTA W11)“ (M. Schmidt, Forschungsinstitut Senckenberg, Frankfurt a. M.)

16.30 Uhr „Zoologie-Daten BIOTA East Africa – Datenmapping in CIDOC CRM“ (K.-H. Lampe, Museum Alexander Koenig, Bonn)

15.7.2005

9.00 Uhr „CIDOC CRM im BIOTA-Kontext – Vorteile und Konsequenzen“ (G. Muche, BIOTA AFRICA, Universität Hamburg)

9.30 Uhr „Vegetationsdatenbanken – Stand der Technik und Ausblick aus mitteleuropäischer Sicht“ (J. Ewald, FH Weihenstephan, Freising)

10.00 Uhr „ESVeg Austauschformat für Vegetationsdatenbanken“ (R. May, BfN, Bonn; P. Strobl, Salzburg)

Die Präsentationen sind als pdf-Dateien auf der Homepage der IT-Fachgruppe hinterlegt ([http://132.180.63.25/GBIF-ITFG/WS\\_Beitraege\\_all.html](http://132.180.63.25/GBIF-ITFG/WS_Beitraege_all.html)).

### ***Verlauf der Diskussion und Ergebnisse***

Herr Muche gab zur Einführung einen kurzen Überblick über GBIF und BIOTA AFRICA und die beiden vorangegangenen Workshops der IT-Fachgruppe. In GBIF-Deutschland werden in der laufenden ersten Phase vor allem Sammlungsdaten digitalisiert und anschließend mittels Datenbank-Wrapper-Technologie im Internet veröffentlicht. Dabei wird angestrebt, dass die Daten bei den bereitstellenden Institutionen verbleiben. Mit Hilfe eines Datenaustauschstandards und eines Wrappers können die Daten aus verschiedensten Datenbanken über ein gemeinsames Portal abgerufen werden. In der laufenden Phase des deutsch-afrikanischen Projektes BIOTA AFRICA werden Daten aus verschiedensten Disziplinen erhoben. Neben biologischen Sammlungs- und Beobachtungsdaten werden auch Daten aus der Sozioökonomie, Fernerkundung, Klimatologie etc. erfasst. Entsprechend den Vorgaben des Projektträgers müssen diese Daten auch nach Abschluss des Projektes nachhaltig verfügbar sein. Darüber hinaus sollen die verschiedenen Daten zwischen Teilprojekten ausgetauscht bzw. zusammengeführt werden können.

Im Mai 2005 fand bereits ein Workshop der IT-Fachgruppe zum Thema „Integration und Anbindung von BIOTA AFRICA Forschungsdaten in das GBIF-Netzwerk“ (Workshop IV, 12. bis 13. Mai 2005, München) statt, hier wurden Datenstrukturen aus verschiedenen BIOTA AFRICA Teilprojekten vorgestellt und die Möglichkeiten und Perspektiven einer Anbindung der BIOTA AFRICA Forschungsdaten an GBIF diskutiert. Auf dem Workshop V („Erstellung eines Datenmodells für die in BIOTA erfassten Standorteigenschaften (Klimatologie, Bodenkunde, Fernerkundung)“, 22. bis 23. Juni 2005, Würzburg) befassten sich die Teilnehmer bereits mit den Datenstrukturen der abiotischen Forschungsbereiche in BIOTA AFRICA (Fernerkundung, Bodenkunde,

Klimadaten) und einem möglichen gemeinsamen Datenmodell, das Herr Weiss bereits vorbereitet hatte.

Dieses vorläufige Datenmodell (*DiversityEcology*) wurde mit der Zielvorgabe entwickelt, die vielfältigen Beobachtungsdaten aus den BIOTA Teilprojekten vereinigen zu können. Als Modul der *Diversity Workbench* könnten dann auch weitere, den Daten entsprechende Komponenten genützt werden.

Es wurde festgestellt, dass das Ziel dieses Workshops darin besteht, das Datenmodell DiversityEcology hinsichtlich der Vegetationsdaten und zoologischen Daten zu prüfen.

### **Das Datenbankmodul DiversityEcology und die Datenstrukturen in vegetationskundlichen und zoologischen BIOTA Teilprojekten**

Herr Weiss stellte das von ihm entwickelte Datenmodell DiversityEcology (als ER-Diagramm) vor. Die unterschiedlichen Datentypen aus dem BIOTA-Projekt sollen in einem gemeinsamen Datenmodell vereint werden können, um gemeinsame Auswertungen zu ermöglichen und gleichzeitig den Datenaustausch zwischen den BIOTA Teilprojekten zu erleichtern. Schließlich müssen die Daten nach Ablauf des Projektes nachhaltig zur Verfügung stehen und öffentlich zugänglich sein. DiversityEcology als Modul der *Diversity Workbench* soll neben ökologischen Daten die Beobachtungen in Raum und Zeit beinhalten (botanisch, zoologisch und abiotisch), alle anderen Daten (Namen, Taxonomie, Literatur etc.) können in andere (größtenteils schon existierende) Module ausgelagert werden.

Herr Weiss erläuterte das Datenmodell in seiner bisherigen Form.

Außerhalb der Optimierung des Datenmodells selbst sind noch folgende Schritte notwendig, bis die Datenbank im Rahmen von BIOTA AFRICA einsatzfähig wäre:

- Evaluation der Austauschstandards ESVeg, GML, EML
- Bestandsaufnahme und Analyse der Daten (soll im Bezug auf botanische und zoologische Daten auf diesem Workshop stattfinden)
- Merkmale, Messmethoden und Wertelisten müssen von den jeweiligen Projektadministratoren definiert werden
- Rechteverwaltung und Logging müssen von den jeweiligen Projektadministratoren festgelegt werden
- BIOTA-intern geklärt werden muss, ob DiversityEcology zentral oder dezentral eingesetzt werden soll
- BIOTA-intern geklärt werden muss, welche Abfragen ermöglicht werden sollen und welche Auswertungsmöglichkeiten und Möglichkeiten der Datenanalyse zur Verfügung stehen sollen

Als Datenbankverwaltungssystem schlug Herr Weiss PostgreSQL vor, da man damit betriebssystemunabhängig bleibt, und diese OpenSource-Software kostengünstiger ist, was auch im Hinblick auf die afrikanischen Projektpartner von Vorteil wäre.

Während der Präsentationen der Datenstrukturen und Inhalte verschiedener vegetationskundlicher und zoologischer Teilprojekte wurden fehlende Bereiche in DiversityEcology ergänzt und Fragen zur Struktur geklärt.

Folgende Punkte im Bezug auf DiversityEcology wurden dabei hauptsächlich diskutiert:

#### Eingabeoberfläche

Auf die Nachfrage nach einer Eingabeoberfläche für DiversityEcology und die Bedienungsfreundlichkeit wurde festgestellt, dass es für DiversityEcology noch keine entsprechende Oberfläche gibt, deren Erstellung (als .Net oder Java Client) relativ aufwändig wäre.

#### Originaldaten

Es wurde darauf hingewiesen, dass bei der Veröffentlichung von Forschungsergebnissen selten die Originaldaten veröffentlicht werden. Es ist im Projekt BIOTA notwendig, diese Originaldaten nachhaltig zu archivieren, so dass diese auch nach Ablauf des Projektes genutzt werden können. Es herrschte Einigkeit darüber, dass die Originaldaten in der Datenbank archiviert werden sollen und keine verlustbehafteten Transformationen geleistet werden sollen, außer zusätzlich zu den Originaldaten (z. B. „berechnete Felder“).

#### Sensible Daten

In vielen BIOTA Teilprojekten werden sensible Daten gesammelt, die nicht ohne weiteres veröffentlicht werden sollen. Geographische Daten, also die Standortinformationen, können auf den verschiedenen Ebenen gesperrt werden, auch in Inventarlisten müssten Namen ausgeblendet werden können.

#### GIS-Funktionalität

Eine performante Bearbeitung raumbezogener projektübergreifender Fragestellungen ist nur möglich, wenn sich die Daten in einer Datenbank befinden. Abfragen über Webservices sind für Einzelfragen und einzelne Datensätze denkbar, aber eine Nutzung von GIS-Funktionalität wird auf diesem Weg wohl kaum zu realisieren sein. Herr Weiss erwähnte als Vorschlag hierzu: PostGIS (stellt GIS-Funktionalität in der Datenbank zur Verfügung, läuft auf Windows und Linux und ist OpenSource)

Die Meinung der Gruppe tendierte in die Richtung, dass der Weiss-Ansatz alle erkennbaren Daten der Vegetationskundler und Zoologen aufnehmen kann, jedoch um den Preis der Abstraktion, so dass schwer abschätzbar ist, wie komplex es sein wird, Auswertungen mit den Daten machen zu können, da die Interpretation der logischen Zusammenhänge in die SQL-Abfragen gesteckt werden muss. Außerdem wurde die Frage aufgeworfen, wenn ohnehin eine Neuentwicklung aufgesetzt wird, warum dann

nicht gleich die Erkenntnisse von Arbeitsgruppen wie der, die CIDOC CRM entwickelt hat, mit aufgenommen werden. Es wäre denkbar, DiversityEcology als Implementierungsform eines konzeptionellen Ansatzes zu nehmen. Hierbei wäre zu überlegen, wie die „Mehrfachvererbungen“ der Objekte überhaupt in einer relationalen Datenhaltung umgesetzt werden können.

#### Sonstiges

Es wurde festgestellt, dass als Standortparameter auch Gewässerparameter in die Datenbank aufgenommen werden müssen, dies ist aber problemlos möglich. DNA Sequenzen können entweder über ein entsprechend angeschlossenes Modul verwaltet, ausgewertet oder zusammengeführt werden, oder über den entsprechenden Link zur kommerziellen DB, bei der die Sequenzen veröffentlicht sind (EMBL, GenBank).

### **CIDOC CRM**

Im Anschluss stellte Herr Lampe CIDOC CRM, ein objektorientiertes, eventzentrisches Begriffsklassenmodell vor, das in Kürze zum ISO-Standard werden soll, und besonders zur strukturellen Überprüfung von Daten geeignet ist.

### **CIDOC CRM im BIOTA-Kontext – Vorteile und Konsequenzen**

Nach einer kurzen Zusammenfassung von Herrn Muche zur Diskussion vom Vortag erläuterte Herr Heinecke die Eigenschaften und Anwendungsmöglichkeiten von CIDOC CRM, und wie eine Modellierung der Daten aus dem BIOTA-Projekt mit CIDOC CRM aussehen würde. Mit CIDOC CRM können vorhandene Datenmodelle überprüft werden, die Integration weiterer Daten erleichtert, sowie Datenaustausch, Datenweitergabe und Implementation erleichtert werden.

Als Vorschlag für die weitere Vorgehensweise nannte Herr Heinecke das Etablieren einer Arbeitsgruppe „CIDOC CRM Modellierung“. Diese Gruppe soll Daten aus einem Teilprojekt mittleren Umfangs und Komplexität mappen, d.h. dem CIDOC CRM Schema entsprechend strukturieren und mit diesem Prototyp („BIOTA-Profil“) die weiteren Domain-Experten interviewen und das Modell so sukzessive vervollständigen.

### **Vegetationsdatenbanken**

Herr Ewald, Herr May und Herr Strobl gaben einen Überblick über Vegetationsdatenbanken und den im Entstehen befindlichen Austauschstandard für Vegetationsdatenbanken ESVeg.

Im Unterschied zu den Forschungsdaten in BIOTA ist es bei Vegetationsuntersuchungen normalerweise üblich, die Primärdaten zu veröffentlichen, so dass diese „nur“ noch digitalisiert werden müssten. In die Vegetationsdatenbank des BfN, Ve-



getWeb, werden in Zukunft die in der Zeitschrift Tuexenia veröffentlichten Primärdaten eingegeben, so dass diese auch digital veröffentlicht sind.

Thema ist derzeit die Integration vorhandener Datenbestände, die z. B. in TurboVeg, VegBank oder VegetWeb vorliegen. Der Weg zum Zusammenbringen unterschiedlich strukturierter Datenbanken wird hier im Aufbau von gemeinsamen Austauschformaten gesehen, also eine Import-Export-Schnittstelle für diese Produkte, die alle Produkte verwenden können. Ein derartiges Austauschformat wird derzeit in XML-Technik erarbeitet.

Während die DiversityEcology-Variante als tendenziell zentrale Datenhaltung eingeschätzt wurde, wurden Austauschformate als Möglichkeit für eine dezentraler Datenhaltung gesehen.

### **Abschlussdiskussion**

In der nachfolgenden Diskussion um die Funktionalität von DiversityEcology und die weitere Vorgehensweise wurde festgestellt, dass DiversityEcology aus botanischer und zoologischer Sicht alle Daten aufnehmen kann, jedoch eine funktionelle Eingabeoberfläche und ein adäquates Austauschformat/Austauschstandard noch erstellt bzw. gefunden werden müssen. Als großer Vorteil des Moduls DiversityEcology wurde die Einbindung in die *Diversity Workbench*, und die Möglichkeit der Nutzung weiterer Module gesehen.

Des Weiteren wurde über die beste Strategie zur Erreichung einer wirklich nachhaltigen Datenhaltung diskutiert. Ein Weg wäre, zunächst alle Daten CIDOC CRM-konform zu strukturieren, was jedoch aufgrund der Vielzahl von Teilprojekten und Forschergruppen und der z. T. schlecht verständlichen Dokumentation von CIDOC CRM nicht praktikabel zu sein scheint. Hier wäre es sinnvoll, eine Expertengruppe zu etablieren, die die Forscher bzw. Projektdatenbankadministratoren bei der Neustrukturierung unterstützt. Ein anderer Weg wäre, im Rahmen von BIOTA AFRICA das Datenbankmodul DiversityEcology als gemeinsame Plattform zu verwenden, Formate und Werkzeuge für den Datentransfer festzulegen bzw. zu entwickeln, sowie einen entsprechenden Datenbank-Client zu entwickeln, so dass DiversityEcology zunächst innerhalb von BIOTA als gemeinsame Basis verwendet werden kann und so die Veröffentlichung der Daten nach Abschluss des Projektes auch über GBIF sichergestellt wird.

Im Plenum wurde besprochen, dass die weitere Vorgehensweise darin bestehen soll, eine Expertengruppe CIDOC CRM zu bilden (Vorschlag: A. Heinecke, G. Mucbe, M. Weiss, K.-H. Lampe), die sich zweimal (evtl. in unterschiedlicher Zusammensetzung) treffen und einen „BIOTA-Prototyp“ anhand eines BIOTA Teilprojektes erstellen soll. Dieser soll dann auf die weiteren Teilprojekte erweitert werden. Schließlich kann das Ergebnis mit der Struktur von DiversityEcology verglichen werden. Für die Bereitstellung an finanziellen Mitteln muss zuvor eine Entscheidung herbeigeführt werden.

## **Protokoll zum Workshop VII der IT-Fachgruppe (GBIF-D)**

### **„Datenbanken für und innerhalb BIOTA AFRICA“**

Beginn:	29. September 2005 13.00 Uhr
Ende:	30. September 2005 13.00 Uhr
Ort:	Zoologisches Forschungsinstitut und Museum Alexander Koenig, Adenauerallee 160, 53113 Bonn
Leitung:	K.-H. Lampe
Protokoll:	G. Muche
Teilnehmer:	Gregor Hagedorn (BBA Berlin), Jürgen Hoppe (Universität Ulm), Karl-Heinz Lampe (Museum Alexander Koenig, Bonn), Gerhard Mu- che (BIOTA AFRICA, Universität Hamburg), Henning Sommer (Uni- versität Bonn), Gerhard Rambold (Universität Bayreuth, ITFG- Koordination)

#### ***Tagesordnung***

29. 9. 2005

- Strategische Ausrichtung
- Auswertung der Workshops in München, Würzburg und Frankfurt
- BIOTA Zielrichtung
- Kriterienkatalog
- Datensicherung, Datenverfügbarkeit
- GBIF-Anbindung

30. 9.2005

- Wissenskartierung der in BIOTA AFRICA involvierten Disziplinen
- Projektanspruch und Realisierungsmöglichkeiten
- Wünschenswerte IT-Strukturen in BIOTA AFRICA für die kommenden Jahre (Empfehlungen)

### ***Verlauf der Diskussion und Ergebnisse***

Entlang der Tagesordnung wurden in sachlicher Diskussion Anspruchsniveau, Umsetzungshemmnisse und Realisierungsszenarien besprochen. Die Kernaussagen lassen sich dabei wie folgt zusammenfassen:

Der BIOTA AFRICA Datenbestand ist mit einem Volumen von ca. einem Terrabyte umfangreich. Die Schwierigkeiten bereitet dabei jedoch weniger das Datenvolumen, sondern die Heterogenität, die durch die Methoden und Tools, die die Disziplinen jeweils benutzen, bedingt ist. Der Datenbestand ist aus diesen Gründen bislang nicht als eine einzige Datenbank zu verwalten, sondern als ein so genannter „Datenpool“, der als Kollektion unterschiedlichster Dateien anzusehen ist. Ein Nahziel des Datenmanagements ist, eine sachlich angemessene Ordnung für den Datenpool aufzubauen (Metadatenbank). Dazu soll geprüft werden, ob eine in Marburg im Rahmen eines DFG-Projekts aufgebaute Metadatenbank an die Bedürfnisse von BIOTA Africa angepasst und genutzt werden kann. Mittelfristig jedoch ist erforderlich, dass eine Verlinkung der Informationen der disziplinären und/oder lokalen Datenbestände auf inhaltlicher Ebene stattfinden kann (Knowledge-Management statt Datenmanagement). Wegen der endlichen Laufzeit des BIOTA AFRICA Projektes (Finanzierung durch das BMBF bis längstens 2009) ist bereits jetzt eine Strategie zu entwickeln, wie der Datenpool von BIOTA AFRICA als „lebende“ Datenbank über das Projektende hinaus zum Nutzen der Wissenschaft erhalten bleiben kann. Dieses wird nur gehen, wenn zwei Bedingungen erfüllt werden. Zum einen muss der Datenbestand bei einer Institution, die als Aufgabe hat, das kulturelle Erbe zu erhalten und deren Existenz deshalb auf Dauer von der Gesellschaft gewährleistet wird („Memory Institution“), untergebracht werden. Als Beispiel wurde das Herbarium Hamburgense und das Zoologische Museum Koenig genannt. Zum anderen muss sichergestellt werden, dass die digital gespeicherten Daten auch entsprechend dem technischen Entwicklungsstand angepasst werden können. Dieses wiederum bedeutet, dass die Strukturen und Inhalte soweit aufbereitet sind, dass sie internationalen Standards entsprechen.

Als ein Umsetzungshemmnis beim Datenbankmanagement wurde gesehen, dass es zunehmende Anforderungen an die Verfügbarkeit der Daten gibt. Publikationen, die auf empirischem Material basieren, müssen nachprüfbar sein. Zu diesem Zweck und weil die Daten kostenintensiv erhoben werden und deshalb weiteren Verwendungen zugänglich sein sollten, müssen die Primärdaten verfügbar sein. Die Projektträger verlangen dieses zunehmend. In der Berliner Erklärung (<http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>) und dem Data sharing protocol (BIOTA AFRICA) haben sich die Wissenschaftler zu diesem Ziel bekannt. In der Praxis ist jedoch Zurückhaltung zu spüren, der begegnet werden muss.

Die ehrgeizigen Ziele sind mit den knappen Ressourcen von BIOTA AFRICA nur schwer zu erreichen, da BIOTA AFRICA sich als Forschungsprojekt versteht und das

Datenmanagement lediglich Servicefunktion innerhalb des Projektes erfüllen kann und nicht etwa Entwicklungsarbeiten grundsätzlicher Art mit IT-Charakter leisten kann. Die Forderung der Nachhaltigkeit und die Bedingtheit der knappen Mittel legen nahe, dass eine möglichst gute Zielerreichung dadurch angestrebt wird, indem mit Institutionen, Initiativen und/oder Netzwerken zusammengearbeitet wird, deren Anliegen mit denen von BIOTA AFRICA überschneidende Bereiche haben, z. B. GBIF, CIDOC CRM, Diversity Workbench oder SysTax.

Das Datenmanagement von BIOTA AFRICA kann, um den Ansprüchen gerecht zu werden, nicht durch eine einzige Person sichergestellt werden. Die Abschätzung ergab, dass für Umsetzungsarbeiten und Serveradministration ein weiterer Mitarbeiter für die Phase III beantragt werden sollte. Der Aufbau eigener Serverkapazitäten wurde sehr empfohlen, um eine der Sachlage angemessene Unabhängigkeit für projektrelevante Aufgaben sicherstellen zu können, da aus anderen Projekten die Unterstützung durch universitäre Rechenzentren als abträglich erfahren wurde. Bedenken zum Aufwand des Betriebes eigener Server (Abwehr von Viren, etc, Firewall) wurden dahingehend entkräftet, dass BIOTA AFRICA Server sicherlich ins Sicherheitskonzept der Universität einzubinden sind, dieses vom Rechenzentrum vermutlich auch verlangt wird und dafür auch Serviceleistung in Anspruch genommen werden kann.

Das Datenmanagement von BIOTA AFRICA enthält nicht lediglich nur die technische Komponente, wie der Datenpool, die Metadatenbanken und die inhaltlichen Datenbanken strukturiert und gesichert werden müssen. Dazu gehört auch die Beschaffung der Daten von den einzelnen Disziplinen. Hierbei gibt es Irritationen bei den Datenproduzenten (Sicherheitsaspekte, Anspruch an die Datenqualität). Ein nicht unwesentlicher Aspekt ist auch die Begrifflichkeit, die zu Sprachbarrieren zwischen den Disziplinen führt. In der Gesprächsrunde entstand der konkrete Vorschlag zum Aufbau einer so genannten „Wikipedia“. Gregor Hagedorn bot an, innerhalb von drei Wochen einen Zugang zu der von ihm konzipierten Wikipedia, in der die BIOTA-Team-Mitglieder interaktiv ihr eigenes Glossar aufbauen können.

## **Protokoll zum Workshop VIII der IT-Fachgruppe (GBIF-D) (zugleich 4. Sitzung der IT-Kommission)**

### **„Perspektiven für den zukünftigen Webauftritt von GBIF Deutschland“**

Beginn:	23. Februar 2006	13.00 Uhr
Ende:	24. Februar 2006	12.00 Uhr
Ort:	Universität Bayreuth, Naturwissenschaften II, Konferenzraum 4, Universitätsstr. 30, 95440 Bayreuth.	
Leitung:	A. Güntsch	
Protokoll:	S. Oehlschläger	
Teilnehmer:	Andreas Allspach, (Forschungsinstitut Senckenberg, Evertebrata III), Anton Güntsch (IT-Kommission, BGBM Berlin), Andreas Hirner (DLR), Joachim Holstein, (Naturkundemuseum Stuttgart, Evertebrata I), Stefan Holzheu (BayCEER, Universität Bayreuth), Alexandra Kehl (IT-Fachgruppe, Universität Bayreuth), Lothar Menner (Forschungsinstitut Senckenberg, Evertebrata III), Gerhard Muche (BIOTA AFRIKA, Hamburg), Andrejs Naumovs (DSMZ, Protisten), Susanne Oehlschläger (GBIF-D, BGBM Berlin), Gerhard Rambold (IT-Fachgruppe, Universität Bayreuth), Dominik Roepert, (BGBM Berlin, Pflanzen (+ Protisten), Michael Wiedemann (SysTax, Universität Ulm).	

#### ***Tagesordnung***

23.2.2006

- |       |  |
|-------|--|
| 13:00 | Begrüßung  |
| 13:30 | Aktueller Stand der Website-Pflege von GBIF-D, Vor- und Nachteile der Nutzung von ZOPE, mögliche Optionen für die Zukunft (A. Güntsch, S. Oehlschlaeger)                                 |
| 14:00 | Bestandsaufnahme der bisherigen Erfahrungen bei der Nutzung von ZOPE bzw. Möglichkeiten des Website-Managements im Rahmen von GBIF-D und zukünftig geplante Inhalte der Sites der GBIF-D |

	Knoten (je ca. 15 min)
17:00	BayCMS – ein php-basiertes Content Management System der Universität Bayreuth (S. Holzheu, BayCEER, Universität Bayreuth)
17:45	Apache Cocoon – eine Option für den Aufbau eines Content Management Systems für GBIF-D? (A. Hirner, DLR)
24.2.2006	
9:00	Abschlussdiskussion zur weiteren Vorgehensweise den Webauftritt GBIF-D betreffend
11:00	Themen der IT-Kommission
12:00	Ende

### ***Verlauf der Diskussion und Ergebnisse***

#### **1. Aktueller Stand der Website-Pflege von GBIF-D**

(Anton Güntsch, Susanne Oehlschlaeger)

Aufgrund reduzierter Mittel für Infrastruktur und Koordination muss der Webauftritt von GBIF-Deutschland neu überdacht werden. Ziel ist es, den Webauftritt mit einem möglichst geringen administrativen Aufwand, einer leichten Pflegbarkeit der Inhalte und einer positiven Darstellung von GBIF-D zu realisieren.

GBIF-Typologie: Mit Ausnahme des Mykologie, Vertebraten und Prokaryonten Knotens mit Satellitenseiten, aber einheitlichem Erscheinungsbild, laufen alle Knoten zentral unter ZOPE in Berlin. Ein Großteil dieser Knotenseiten wird jedoch gegenwärtig nur wenig aktualisiert oder erweitert.

Die Hauptseite von GBIF-D wird mit ZOPE in Berlin gemanagt. Die GBIF-D Homepage informiert über Hintergründe und Projektstruktur, gibt Veranstaltungshinweise, offeriert Materialien und eine Statistik zu den online Daten und gibt Kurzbeschreibungen und Verlinkung von Teilprojekten.

Zugriffsstatistik: ca. 75.000 Zugriffe auf die GBIF-D Webseite (ohne Zugriffe von Webcrawlern); Spitzen wurden im Mai 2005 registriert, die meisten Zugriffe kamen unerklärlicherweise aus Japan.

Was ist ZOPE?

ZOPE = **Z** Object **P**ublishing **E**nvironment

- Open-Source-Webanwendungs-Server
- ZOPE ist Content Management System
- ZOPE ist Application-Server
- Implementiert in Python und C
- Daten werden in der mitgelieferten ZODB abgelegt
- Eigene Skriptsprache DTML
- Eigener Web-Server

Vorteil	Nachteil
<ul style="list-style-type: none"> <li>• Hohe Zuverlässigkeit</li> <li>• Relativ leicht zu administrieren</li> <li>• Mächtige Versions-Verwaltung</li> <li>• Große Verbreitung</li> <li>• Zahlreiche Module verfügbar</li> </ul>	<ul style="list-style-type: none"> <li>• Kein WYSIWYG-fähiger Editor</li> <li>• Voransicht von Seiten erst nach Publikation möglich</li> <li>• Schulungsaufwand (html-Kenntnisse nötig)</li> </ul>

→Die zukünftige Plattform könnte wie folgt aussehen: Variante 1: Alternatives CMS; Variante 2: HTML ohne CMS; Variante 3: ZOPE.

Um die Attraktivität zu erhöhen sollte die Webseite aktueller wirken, d.h. News auch vom IT-Bereich sollten prominent auf der Homepage stehen (wie bei <http://www.biocase.org> und <http://www.gbif.org>). Zusätzlich könnte dieses News-system mit einem RSS-Feed ausgestattet werden, so dass GBIF-D Nutzer die Nachrichten abonnieren können.

## 2. Bestandsaufnahme der bisherigen Erfahrungen und zukünftig geplante Inhalte der Sites der GBIF-D Knoten (je ca. 15 min)

**Lothar Menner, Evertebrata III:** System: ZOPE. Der Inhalt der Webseite wurde mangels Bedarfsanmeldung durch die Teilprojekte kaum erweitert oder verändert. Es fehlte eine Zielsetzung sowie eine Person für die Pflege der Webseite. Zudem ist ZOPE für viele nicht nutzbar, da HTML-Kenntnisse erforderlich sind. Inhalt: Ziele und Strategien des Knotens, Aufzählung der Teilprojekte.

→Die GBIF Webseite wird von den Projektteilnehmern kaum als Informationsquelle genutzt. Es fehlen direkte Verknüpfungen von der Hauptseite zur Datenbank bzw. Suchabfrage.

**G. Rambold:** Bei GBIF International wird die Datenabfrage zentral abgedeckt. Die Teilknoten von GBIF-D sollten jedoch erhalten bleiben und die politische Verantwortung für die Inhalte übernehmen. Die Knoten werden auf Landesebene gefördert. Daher ist eine zentrale Suchabfrage zwar eine elegante Lösung, jedoch politisch nicht durchsetzbar.

**Alexandra Kehl für Mykologie:** System: Statische HTML-Seiten auf Webserver, Seiten werden in München (BSM) verwaltet. Die Verwaltung läuft in CVS, d.h. mehrere Personen können an der Webseite arbeiten. Ein Editieren erfolgt über den HTML-Editor. Inhalt: Reine Portalfunktion, Zugriff auf verschiedene Verknüpfungen und Publikationen.

**Alexandra Kehl, IT-Fachgruppe:** Statische HTML Seiten auf Webserver (Orientierung an Mykologie-Seite); Website wird von der IT-Fachgruppen-Koordination in Bayreuth verwaltet. Inhalt: 1) Workshopankündigung und Dokumentation, Personenverzeichnis 2) Dokumentation der IT-Fachgruppe, Online-Formulare. Integration der Datenbank aus München (Übernahme von LIAS), Online-Formulare: Einloggen mit Passwort z.B. für Ergänzungen von bereits ausgefüllten Formularen; Änderungen werden per Hand eingegeben!

→Der Link zur IT-Fachgruppe auf der GBIF-D Website sollte besser zu finden sein

→BayCMS ist leichter zu handhaben als HTML

Anton Güntsch: Endprotokoll des IT-Fachgruppen-Surveys wäre gut.

**Joachim Holstein, Evertabrata I:** System: ZOPE, kommt damit gut zurecht. Bezüglich des Inhalts der Webseite wurden Teilprojekte oft kontaktiert, jedoch gab es kaum Rücklauf. Inhalt: Aufzählung der Teilprojekte mit Links, Datenabfrage erfolgt über SysTax (Link). Der Knoten wird mit Institutsmitteln auf einer niedrigen Ebene gepflegt.

→Angleichung an GBIF-International begrüßenswert

**Michael Wiedemann, Vertebrata:** System: ZOPE: Pflege der Webseite ist nicht auf neuestem Stand. Inhalt: Ziele und Hinweise zur Datenabfrage. Links zur Datenabfrage (Ulm), Taxonbrowser und Zoologieportal (liefert alle Daten Deutschlands, Aufbau eines zentralen Index-Systems für schnelle Zugriffe), Layout erfolgt über Templates, SysTax verwendet Perlsripts und kann php nicht unterstützen.

→Frage: Integration von Primärinhalten?



**Andrejs Naumovs, Prokarya:** System: eigenes CMS, HTML Basis, php script; CMS wird seit drei Jahren erfolgreich genutzt und ist einfach zu handhaben. Datenbank: SQL, Apache, php. Suchabfrage wurde nicht auf der Homepage implementiert.

**G. Muche, BIOTA AFRICA:** Bei BIOTA AFRICA war eine Arbeitsteilung notwendig, daher wird für die BIOTA AFRICA-Webseite (<http://www.biota-africa.com/800/start.htm>) kein CMS genutzt. Die BIOTA-Leute bauen ihre Publikationen in die Publikationsliste der Webseite eigenständig ein. Aufgrund der Darstellung der eigenen Leistung sowie die Aussicht auf eine dritte Phase von BIOTA AFRICA, ist der Rücklauf sehr hoch und die Effizienz dieser Liste sehr groß. Zu Beginn standen 500 Publikationen zur Verfügung, nach drei Monaten waren es 1000! Mit der Veröffentlichung der Primärdaten im Netz sind die Leute eher zögerlich.

Die Publikationszuordnung erfolgt projektbezogen, genannt werden Autor etc., Abstracts, Keywords, Word- und pdf-Dateien sind einladbar, eine Statistik ist ebenfalls abrufbar. Die Eingabe erfolgt mit Passwort, dennoch konnten viele Falscheintragungen registriert werden.

Administration: Insgesamt haben 97 Leute Zugriffsrechte, allerdings nutzen weniger als die Hälfte die Webseite aktiv.

→Eine Statistik (Welche? Eine rudimentäre Statistik-Seite gibt es bereits) wäre auch für GBIF-D interessant

→Wäre solch ein „Wettbewerb“ auch bei GBIF-D möglich?

### **3. BayCMS – ein php-basiertes Content Management System der Universität Bayreuth (Stefan Holzheu)**

Die Webseite des Bayreuther Zentrums für Ökologie und Umweltforschung (Bay-CEER) basiert auf PostgreSQL und PHP. Es handelt sich um eine relationale Datenbank mit Webfrontend. Die Anwendung des Systems läuft bei ca. 20 Lehrstühlen und Abteilungen, diversen Forschergruppen, SFB481, Elitenetzwerk, Tagungsanmeldung.

Aufbau der Webseite

1995 Aufbau der Datenbank (Verwaltung und Messdaten) in Oracle.

1998 Webseitenaufbau aus Datenbank mit Perl.

2002 Umstellung auf PostgreSQL und PHP.

## Anforderungen

- Ablage strukturierter Inhalt→Content Objekte
- Einfache Datenpflege→Formulare
- Zuordnung von Objekten zu unterschiedlichen Organisationen, d.h. die Daten werden einmal abgelegt und können in verschiedenen Zusammenhängen verschieden dargestellt werden.
- Individuelle Darstellung der Organisationen über Templates.
- Benutzerverwaltung

## Stärken und Schwächen des Systems:

Vorteil	Nachteil
Verwaltung der Struktur und Inhalte Anordnung der Inhalte Flexibel einsetzbar	Freie Inhalte schwer einbaubar Sicherheitslücken Version-Control nur rudimentär

## PostgreSQL:

Vorteil	Nachteil
Open source Referentielle Integrität Prozedurale Sprachen	

## PHP:

Vorteil	Nachteil
Schnelle Entwicklung Schnelle Ausführung Flexibel	Häufig unübersichtliche Programmierung

## Ausblick

- Webseitenpflege: Unterteilung in Verwaltung (reine Verwaltung) und Administration (tiefere Beschäftigung mit System)→ein zukünftiger Umstieg auf ein offenes System ist denkbar.
- Normale Nutzer können das System ohne Probleme nutzen.
- Gestaltung der freien Seiten erweist sich allerdings als schwierig.

- Nutzung eines internen HTML-Editors in CMS möglich.

→ BayCMS bietet den Knoten/GBIF-D Hilfestellung bei BayCMS an. Allerdings müssen die Inhalte klar formuliert sein. Bayreuth könnte auch die GBIF-D Webseite hosten.

#### **4. Apache Cocoon – eine Option für den Aufbau eines Content Management Systems für GBIF-D? (Andreas Hirner)**

Sino-German Coal Fire Research, Projekt des DLR  
([http://www.coalfire.caf.dlr.de/intro\\_en.html](http://www.coalfire.caf.dlr.de/intro_en.html))

Webseite: Publishing Frame Work Servlet in php und XML (zu 100 % in Java realisiert), modular aufgebaut

Ziel:

- Metadatenverwaltung vieler internationaler Projekte
- Entwicklung einer modularen Web-Applikation
- Betriebssystem-Unabhängigkeit
- Ansprechendes Layout/Design
- Sicherheit
- Schnelle Anpassung der Datenstruktur

Warum Apache Cocoon?

- System wurde schon lange genutzt und hat gute Ergebnisse erzielt
- Internationalisierung (Sprache)
- Nutzung einer einfachen Datenbank
- Einbindung von Dateien, externen Webseiten möglich
- Dokumente werden als XML abgelegt
- Datenbanken können verknüpft werden, Suchmaschine
- Module müssen selbst programmiert werden → Java Kenntnisse sind nötig
- Formulare können eigenständig delegiert und validiert werden

→ Für eine Neuentwicklung ist das System „genial“, weil man es eigenständig und nach seinen eigenen Maßstäben entwickeln kann.

Herzstück = Sitemapdatei:

- Aufruf von Java Script Controller
- Einbindung von Driftprogrammen
- Einbau eigener Werkzeuge

Apache Cocoon = komplexe Java application: XML, Java, XSCT, Javascript (kein CMS)

→hohe Flexibilität

Apache Lenya open source Content Management (<http://lenya.apache.org/>)

Bei den am DLR entwickelten Portalen geht man davon aus, dass nur wenige Nutzer die Inhalte wirklich pflegen. Für diese ist es zumutbar, sogar XML-Dateien zu editieren. Eine spezielle Nutzeroberfläche ist daher nicht implementiert worden.

Freitag, 24. Februar 2006

## **5. Abschlussdiskussion zur weiteren Vorgehensweise den Webauftritt GBIF-D betreffend**

Die Diskussion erbrachte folgenden Empfehlungen der IT-Kommission/Fachgruppe bezüglich der zukünftigen inhaltlichen Gestaltung der GBIF-D Websites:

- 1) Wenn Ferneditierung, dann sollte möglichst ein WYSIWYG-Editor verfügbar gemacht werden.
- 2) Ergebnisse der IT-Fachgruppe sollten mehr ins Zentrum der Webseite gerückt werden.
- 3) Die Ergebnisse der deutschen Biodiversitätsinformatik insgesamt sollten prominenter abgebildet werden.
- 4) Die GBIF-D Website sollte auf ihrer Eingangsseite ein datenbankgestütztes News-System mit RSS-Feed bekommen (vgl. BioCASE).
- 5) Dokumentation der bisherigen Mobilisierung von Daten (tabellarisch und grafisch)
- 6) Illustration des GBIF-D Rahmenpapiers (als Darstellung von GBIF, Förderung, Positionspapier)→Adressat: BMBF

- 7) Zusammenstellung der Sammlungsdaten und Projekt-Metadaten von GBIF-Deutschland zur Aufbereitung als Datenbank, d.h. Erstellen einer Datenbank bzw. eines dynamischen Webinterface mit Schnittstellen zu anderen Sammlungsdaten/Metadaten → Harmonisierung: Aufgabenverteilung zur Zusammenstellung der Daten: BIOTA: G. Muche, BfN: J. Holstein; GBIF-D: A. Güntsch, S. Oehlschlaeger; Rest: G. Rambold, A. Hirner; IBOS-D: J. Holstein schickt Mail mit Daten zu den beteiligten Personen und Institutionen
- 8) Graphische Darstellung von geographischen Daten in den Metadaten (vergleichbar zu [http://www.biocase.org/whats\\_biocase/providers\\_map.shtml](http://www.biocase.org/whats_biocase/providers_map.shtml))
- 9) Orientierung an „look and feel“ der Website von GBIF International
- 10) Anwendung der GBIF-D Daten als Demonstrationsprojekt

→ Weitere Ideen zur Gestaltung der GBIF-D Webseite sollen bis zum 31. August an A. Güntsch gehen.

Deutschland hat einen Nachholbedarf bei der Digitalisierung von Sammlungsdaten. Hier könnten Anträge gestellt werden bei NSF und DFG → Nachforschung

→ Aufruf an GBIF-D Koordinatoren: Daten, die im Rahmen von GBIF-D erfasst worden sind müssen verfügbar und suchbar gemacht werden.

Ergebnisse bezüglich der zukünftigen technologischen Plattform des GBIF-D Portals: Gegenwärtig werden in den verschiedenen GBIF-D Knoten unter anderem die folgenden Entwicklungs-Technologien eingesetzt:

BGBM:	Python, Java, Microsoft, ASP, ColdFusion, JSP
SysTax, Ulm:	Perl, Python
DSMZ:	PHP
Ev I, II, Vertebrata:	SysTax
Mykologie:	ColdFusion, Java Server Pages
Evertebrata III	ASP, PHP

Es ist nicht anzunehmen, dass sich diese Diversität reduzieren wird und ein Trend bzgl. der im GBIF-D Portal einzusetzenden Techniken abzulesen sein wird. Daher soll nun zunächst so vorgegangen werden:

→ Berlin und Bayreuth erstellen bis Ende April eine Gegenüberstellung von ZOPE und dem in Bayreuth entwickelten CMS. Die IT-Fachgruppe und die IT-Kommission bekommen diese Zusammenstellung als Entscheidungshilfe zugeschickt. Die Ab-

stimmung bzw. endgültige Version wird dann den GBIF-D Koordinatoren zur Abstimmung vorgelegt.

→ Alle GBIF-D IT-Zentren sind eingeladen bis Ende Juni ein Konzept für die zukünftige technische Plattform des GBIF-D Portals anzubieten (eingesetzte Techniken, technischer Support, Nutzer-Support, Konzept für Pflege etc.). Die Angebote werden den Koordinatoren zur Entscheidung vorgelegt.

## 6. Aktivitäten bei GBIF International (Anton Güntsch)

Die GBIF Data Portal Strategy nebst diverser Wiki-Pages für Entwickler sind öffentlich. Beiträge sind grundsätzlich erwünscht. Ein Dokument zur „Data Portal Strategy“ (Donald Hobern) befindet sich im Netz.

- Protokolle: DIGIR, BioCAsE, Tapir (= neue USA/Europa Version, sehr leistungsfähig, XML Struktur; schon in der Anfrage kann erfragt werden, welche Daten angefordert werden sollen), Provider-Software ist bereits verfügbar. DIGIR und BioCAsE werden auch zukünftig unterstützt.
- Zukünftig wird der Indexvorgang vermehrt von den Daten Providern gesteuert. Sie entscheiden also, wann und wie oft ihre Daten bei GBIF international aktualisiert werden.
- Schema Repository: Verwaltung von relevanten XML Schema Konzepten, Umstieg auf RDF möglich (Vorteil: rdf ist ohne Versionswechsel erweiterbar und man kann beliebige Beziehungen zwischen Konzepten herstellen und wieder verwenden).
- Validation Chain: Projekt/Überprüfungsmechanismen/ Datenqualitätsüberprüfung im Aufbau. Tools dazu gibt es in Brasilien bei CRIA, „Brazil Centro des Referência em Informação Ambiental“, das ein neues Datenvalidierungs-Programm (<http://splink.cria.org.br/dc>) entwickelte, welches als Web-Service verfügbar ist.
- Master Data Store: Sammler für alle Daten, exakte Kopie der Datenbank. Kopien an Slave Data Stores (Mirror am BGBM, zukünftig noch Korea und USA)

### **Aufgaben und Termine**

- Berlin und Bayreuth erstellen bis Ende April 2006 eine Gegenüberstellung von ZOPE und BayCMS/psp für die IT-Fachgruppe und die IT-Kommission. Die IT-Fachgruppe und die IT-Kommission bekommen diese Zusammenstellung als Entscheidungshilfe zugeschickt. Die Abstimmung bzw. endgültige Version wird dann den GBIF-D Koordinatoren zur Abstimmung vorgelegt.

- Angebote für zukünftige GBIF-D Portal Plattform bis Ende Juni 2006 an Koordinatoren.
- Ideen zur Gestaltung der GBIF-D Webseite bis Ende August 2006 an A. Güntsch.
- Zusammenstellung der Sammlungsdaten und Projekt-Metadaten von GBIF-Deutschland zur Aufbereitung als Datenbank bis spätestens Mai 2006: BIO-TA: G. Muche, BfN: J. Holstein; GBIF-D: A. Güntsch, S. Oehlschlaeger, Rest: G. Rambold, S. Holzheu.

## **Protokoll zum Workshop IX der IT-Fachgruppe (GBIF-D)**

### **"Einsatzmöglichkeiten von DiversityResources zur Verwaltung wissenschaftlicher Bildsammlungen"**

Beginn:	30. November 2006	10.00 Uhr
Ende:	1. Dezember 2006	18.00 Uhr
Ort:	Botanische Staatssammlung München, Menzinger Straße 67, 80638 München. Lesesaal der Bibliothek, EG, Arbeitsräume des IT-Zentrums SNSB.	
Leitung:	Dr. D. Triebel	
Protokoll:	J. Ingenhaag, M. Weiss	
Teilnehmer:	Andreas Beck (BSM, München), Jan Ingenhaag (BSM, München), Andreas Kohlbecker (BBA, Berlin), Asif Lakehsar (BSM, München), Dieter Neubacher (München), Dagmar Triebel (BSM, München), Markus Weiss (BSM, München)	

#### ***Tagesordnung***

1. Begrüßung (D. Triebel)
2. Bildsammlungen im Rahmen der Projekte "BIOTA", "SEBio" und "Poelt & Jahn 1963" (J. Ingenhaag)
3. Einführung in DiversityResources (A. Kohlbecker)
4. Die Online-Bilddatenbank "Fungal Illustrations of TUB" (D. Triebel)

#### **Diskussionspunkte**

- Haupteinsatzfelder von DiversityResources
- Erfassung der publizierten Bildsammlung "Poelt & Jahn 1963"
- Mögliche Interaktionen/Verknüpfungen zwischen NaviKey und DiversityResources, Einbindung von Bildquellen aus dem Internet in NaviKey
- Eignung von DiversityResources für die Projekte "BIOTA" und "SEBio"



- Technische Rahmenbedingungen von DiversityResources
- Massenimporte lokaler Bildbestände und Bildern aus dem Internet
- Ist eine Vorkonfiguration des Webinterfaces und damit eine Anpassung an verschiedene Projekte möglich?
- Präsentation der einzelnen Sammlungen mit Logo und klare Trennung von der Diversity Workbench

30.11.06 (nachmittags) bis 1.12.06: Fortsetzung der Diskussionen, Software-Installationen etc., Abschlussbesprechung

Donnerstag, 30.11.2006

**1. Begrüßung und Organisatorisches durch Dagmar Triebel**

**2. Vorstellung der Tafeln zu mitteleuropäischen Pilzen von Poelt & Jahn 1963 durch Jan Ingenhaag**

- Das Werk von Poelt & Jahn umfasst ca. 160 Tafeln mit einem oder mehreren Organismen pro Tafel auf der Vorderseite und entsprechenden Beschreibungen auf der Rückseite.
- In DiversityResources (BBA) existiert bereits eine Bildsammlung mit Fragmenten, die Darstellung von Fragmenten wird aber von der Weboberfläche von DiversityResources nicht speziell unterstützt. Fragmente und gesamte Abbildung werden gemeinsam hintereinander angezeigt.

Diskussion

- Wenn mehrere Organismen auf einer Tafel vorhanden sind, sollten Bild- und Textinformationen der jeweiligen Organismen dieser Tafel zusätzlich in einzelne Bilddateien zerlegt abgelegt werden und über eine FragmentID der Originaltafel zugeordnet werden.
- Eine Identifizierung der Fragmente wäre über relative Positionsangaben im Übersichtsbild möglich (einfache Rechtecke oder auch Polygone). Relative Angaben sind deshalb von Vorteil, da diese leicht auf alle Qualitätsstufen übertragen werden können

- Eine intensivere Nutzung von Positionsangaben in Bildern kann nur über ein noch zu entwickelndes Plugin realisiert werden, echte GIS-Funktionalität wäre nur über eine PostgreSQL-Datenbank zu erreichen.
- Eine Volltextindizierung ist bisher noch nicht möglich.
- Taxa werden im Augenblick über eine eigene Pickliste abgebildet. Eine Verknüpfung zu DiversityTaxonNames wäre über Keywords mit der Kategorie DiversityTaxonNamesURI denkbar.

### 3. Vorstellung von SEBio durch Jan Ingenhaag

- Bei der Übersichtsdarstellung eines Relevés werden viele Thumbnails in einem Raster angezeigt.
- Zukünftig soll ein zusammengesetztes Gesamtbild unter der Verwendung relativer Positionsangaben verwendet werden (→ Fragmente)
- SEBio soll über die Bildverwaltung hinaus folgende Aufgaben erfüllen:
  - Projektverwaltung
  - Speicherung geographischer Positionsangaben
  - Verwaltung von Objekten in den Relevébildern (z.B. Flechten) und Zuordnung dieser Objekte zu einem taxonomischen Thesaurus
  - Analyse und Berechnungen anhand der identifizierten Bildobjekte
- Für SEBio ist wegen der bereits vorhandenen GIS-Funktionalität nur PostgreSQL geeignet.

#### Diskussion

- Bei der Verwendung von DiversityResources würden zahlreiche Redundanzen in den Metadaten entstehen. Eventuell könnte es aber nur für die Bildverwaltung genutzt werden.
- Evtl. Umsetzung von DiversityResources auf PostgreSQL wegen GIS-Funktionalität
- Evtl. wäre ein Ausbau von DiversityDescriptions als ein Modul für Descriptions/Ecology/SEBio sinnvoll, da alle GIS-Funktionalität benötigen.
- Für Analysen sind weitere Layer notwendig. Eine Beschreibung der Methodik könnte in DiversityDescriptions erfolgen.

- Für SeBio wären Teile von DiversityDescriptions, DiversityResources und DiversityTaxonNames nötig
- In DiversityResources müssten für die Ein- und Ausgabe Plugins entwickelt werden

#### 4. Vorstellung von DiversityResources durch Andreas Kohlbecker

- Vorstellung von Eingangsseite und Beispiele:
  - Suchfunktion
  - Deskriptoren
  - Editiermodus – einfach und advanced
  - Online-Hilfe im Wiki
- Metadaten haben keine Beziehung untereinander
- Deskriptoren können miteinander in Beziehung gesetzt werden. Sie können für eine Collection (entspricht ungefähr „Projects“ in DiversityCollection) eingeschränkt werden.
- Zwischen Deskriptoren können Beziehungen hergestellt werden. Bei Beziehungen zwischen zu beschreibenden Bildobjekten innerhalb von ResourceInstances muss man sich beispielsweise mit Indizes behelfen. Beispiel:
  - Wirtspflanze 1 – Parasit 1 auf Wirtspflanze 1
  - Wirtspflanze 1 – Parasit 2 auf Wirtspflanze 1
  - usw ...
- ResourceInstances können untereinander in eine Parent-Child Beziehung gesetzt werden, wobei das Kindelement ein Fragment des Elternelements ist.
- Demonstration des Anlegens einer neuen Bildsammlung („Collection“)
- Das Einlesen von Bildern ist bislang nur über den JSP-Webclient möglich. Es können manuell max. 50 Bilder gleichzeitig hochgeladen werden. Große Bilddatenbestände wurden bisher durch Gregor Hagedorn direkt über die Datenbank eingespeist.
- Vorstellung des Instance Robot → diverse Funktionen, z. B. Prüfung der Verfügbarkeit von Bildern, automatische Umrechnung der verschiedenen Bildqualitäten

- Die Freigabe einer Collection erfolgt über das ReleaseDate. Bilder, die nach dem EditorialRelease hochgeladen wurden, müssen zusätzlich freigeschaltet werden.
- Die Ablage der Bilder kann auf einem beliebigen Server erfolgen, sofern über Windows-Netzwerk erreichbar oder über URI.
- Thumbnails werden lokal abgelegt
- spezielle Syntax zum Verweis auf Bilder und zum Einbetten derer in andere Webseiten

#### Diskussion

- Ein Mechanismus zum massenhaften Import von im Internet verfügbaren Bildern ist über eine Skriptsprache denkbar (z.B. Webl)
- Auslagerung der Metadaten evtl. überflüssig (?)

### **5. Vorstellung von NaviKey durch Jan Ingenhaag und Dieter Neubacher**

- NaviKey wurde in Harvard zur Prozessierung von Delta-Daten entwickelt.
- Der Quellcode wurde vor 3 Jahren übernommen und inzwischen vollständig umgeschrieben.
- NaviKey greift auf Delta-files zu. Im DiversityNavigator werden diese aus PostgreSQL-DB-Tabellen erzeugt und an NaviKey übergeben.
- IntKey arbeitet auch mit DELTA-Daten, aber nur lokal, nicht online.
- in DEEMY: Vorstellung von NaviKey in Anwendung, Beschreibungen und Bildern.

#### Diskussion

- Es wäre wünschenswert, wenn Bilder, die Characters, CharacterStates oder Items zugeordnet sind, über NaviKey aufgerufen werden könnten. Eine Option mag im Einsatz von DiversityResources liegen.

### **6. Vorstellung der BildDB TUB durch Dagmar Triebel**

- Die Bildsammlung umfasst 1700 Zeichnungen und 1400 EM-Bilder von 530 Pilzen

- Insgesamt ca. 2100 Datensätze mit Metadaten zu den Bildern wie Kurzbeschreibung, Bild-Autor, Maßstab, Belegnummer. Diese Metadaten könnten direkt in DiversityResources übernommen werden.
- Das bisherige Webinterface umfasst auch einen TaxonomyBrowser und eine taxonomische Suchfunktion. Diese können mit DiversityTaxonNames umgesetzt werden

### ***Verlauf der Abschlussdiskussion und Ergebnisse***

- Eine Anpassung der Oberfläche von DiversityResources an einzelne Collections/Projekte wäre wünschenswert.
- Die Standardsuchmaske sollte konfigurierbar gemacht werden, um einzelne Elemente ausblenden zu können
- Im zentralen Konfigurationsfile (Workbench.config) könnte eine Liste anzuzeigender Felder festgelegt werden → dies ist bisher nur pro Installation möglich.
- Evtl. sollten jeder Collection ein eigenes Config-File zugeordnet werden.
- Die Suchmaske sollte auf je nur eine Collection beschränkt werden können.
- Auch für DiversityReferences soll pro Sammlung/Projekt ein eigenes Layout und die Auswahl angezeigter Felder möglich sein

Am Nachmittag des 30.11. wurde mit der Installation der in Berlin entwickelten Komponenten der Diversity Workbench in München begonnen (DiversityUsers). Hierbei wurden für einige Teilbereiche notwendige Anpassungen durch Andreas Kohlbecker vorgenommen. Die noch nicht dokumentierten Teile dieser Anpassungen werden in der Folge in das Diversity Workbench Wiki aufgenommen werden und Teile der Module werden für eine vereinfachte Installation umgeschrieben.

### Freitag, 01.12.2006

Weiterführung der Installation der Diversity Workbench Module.

Abschlussbesprechung Freitag

- Die Installation in München sollte zwei Collections enthalten:
  - Erysiphales – U. Braun 1987
  - Sammlung H. & H. Doppelbaur

- Ein Pilotprojekt in München sollte NaviKey-LIASlight mit Flechtenbildern aus dem Internet verknüpfen
- Dazu wird ein Massenimport über Stored Procedures direkt über die Datenbank nötig, bei dem gleichzeitig Metadaten gesetzt werden
- Welche Metadaten?
  - ImageURI
  - Titel
  - Bildautor
  - Ort
  - ...
- Massenimport über CSV-Datei (tab-separiert, UTF-8)  
1. Zeile mit Filename/Feldnamen/Descriptoren  
Format der CSV ist zu definieren (→ Andreas Kohlbecker)
- Beispiele für CSV-Format:
  - FileName, SourceDescription, OriginalOrder, DescriptorType1, DescriptorContent1, ...
  - FileName, SourceDescription, OriginalOrder, MetadataType, MetadataContent, ...

Im Anschluss an den Workshop werden durch Andreas Kohlbecker die im Laufe der Installation aufgetretenen Schwierigkeiten behoben und die vorgeschlagenen Verbesserungen in die Module aufgenommen.

## **Protokoll zum Workshop X der IT-Fachgruppe (GBIF-Deutschland):**

### **Datenverwaltung im Rahmen der Forschungsaktivitäten in den DFG-Exploratorien**

<b>Beginn:</b>	14. Juni 2007	13.00 Uhr
<b>Ende:</b>	15. Juni 2007	13.00 Uhr
<b>Ort:</b>	Universität Bayreuth	
<b>Leitung:</b>	Gerhard Rambold	
<b>Protokoll:</b>	Alexandra Kehl, Simone Pfeiffer	
<b>Teilnehmer:</b>	Oliver Archner (BayCEER, Bayreuth), Calin Arens (MPI, Jena), Dennis Heimann (MPI, Jena), Jörg Holetschek (BGBM, Berlin), Alexandra Kehl (Uni Bayreuth, GBIF IT-Fachgruppe), Reiner Krause (MPI Jena), Jens Nieschulze (MPI, Jena), Simone Pfeiffer (Uni Potsdam), Gerhard Rambold (Uni Bayreuth, GBIF IT-Fachgruppe), Marco Schmidt (Forschungsinstitut Senckenberg, Frankfurt), Bernhard Volz (Uni Bayreuth, Angewandte Informatik IV), Markus Weiss (IT-Zentrum der SNSB, München).	

#### **Zeitplan**

##### **Donnerstag, 14.06.2007**

13.30 Uhr Begrüßung (Rambold)

Kurze Vorstellung der IT-Fachgruppe (Kehl, Rambold)

13.45 Uhr Bedarfsanalyse bzgl. Daten und Datenhaltung im Rahmen der geplanten Forschungsaktivitäten in den DFG-Exploratorien (Nieschulze)

Bereits bestehende Datenhaltungssysteme und IT-Entwicklungen am MPI Jena (Krause)

Die Messungen-Datenbank am BayCEER, Status quo und zukünftige Erweiterungen (Archner)

16.00 Uhr Pause

- 16.30 Uhr Die Komponenten der Diversity Workbench: Gesamtkonzept, Modelle, Implementierungen (Weiss)
- Austauschformate für primäre Biodiversitätsdaten: TDWG Datenstandards und Protokolle (Holetschek)
- Prozessmanagement für die Integration und Analyse von in heterogenen Datenbanken gespeicherten biologischen Forschungsdaten (Volz)
- 19.45 Uhr Abendessen

## **Freitag, 15.06.2005**

- 09.00 Uhr Weiterführende Diskussion

---

### ***Begrüßung durch Gerhard Rambold***

#### **Vorbemerkungen zum Workshop X**

#### **Erfahrungen zum Datenmanagement in Biodiversitäts-Projekten**

Beim Beginn von Projekten im Bereich der Biodiversitätsforschung stehen beteiligte Institutionen bzw. Arbeitsgruppen häufig vor der Frage ob, (a) bereits vorhandene Datenhaltesysteme und damit operierende Applikationen genutzt und den neuen Fragestellungen angepasst, (b) neue Datenbanken und Anwendungen entwickelt, oder aber (c) andernorts bereits existierende Systeme in das eigene integriert werden sollten.

Speziell bei kleineren, stärker spezialisierten Arbeitsgruppen sind die vorhandenen IT-Strukturen oft Einzelplatzlösungen, mittels derer anfallende Biodiversitäts- und Umweltdaten zwar strukturiert, aber nicht durchgängig standardisiert vorgehalten werden. Dies geschieht z. B. anhand von Tabellenkalkulationsprogrammen oder gar bloßen Textdateien, die oft nur nach den Vorstellungen einzelner Projektmitarbeiter (Diplomanden, Doktoranden) entwickelt wurden. Zumeist fehlen auch die technischen Möglichkeiten, um Intranet- oder Internet-basiert Eingaben in verteilte oder zentrale Datenbanken vornehmen zu können.

Jedoch erzeugen auch auf speziellere Fragestellungen fokussierte Arbeitsgruppen nicht nur Daten einer einzigen Datendomaine (z. B. georeferenzierte Beobachtungsdaten, Messreihen-Daten, pflanzensoziologische Daten) sondern eine Kombination derselben. Diese werden oft gemeinsam in mehr oder weniger ausreichend dokumentierten Datenhaltesystemen eingepflegt und höchst selten in nach Datendomai-



nen eigenständigen, bzgl. ihres Aufbaus und der verwendeten Ontologien gut dokumentierten Datenbanken verwaltet. Die verwendeten Daten(-typen) folgen nur selten einem von einem internationalen Gremium empfohlenen Standard und die in Einzelprojekten anfallenden Daten sind daher in den seltensten Fällen problemlos miteinander kombinierbar bzw. korrelierbar.

Im Falle eines tatsächlich stattfindendem Datenaustausches zwischen einzelnen Arbeitsgruppen im Rahmen angestrebter gemeinsamer Publikationsarbeit findet dieser zumeist auf traditionellem Wege statt, d. h. also nicht automatisiert zwischen interoperablen Datenbanken oder via Webservices.

Unter den geschilderten Bedingungen, d. h. bei Fehlen ausreichender Datenstrukturierung, -standardisierung und -vernetzung, besteht für die langfristige Datenhaltung, -pflege und -bereitstellung das Problem, dass die lokale inhaltliche wie technische Pflege oft nicht ohne Weiteres über Jahre hinweg gewährleistet werden kann und speziell die Übergabe von Daten an neue Mitarbeiter problematisch wird. Erschwerend kommt bei Nichtbeachtung internationaler Datenstandards (soweit vorhanden) das Problem einer mangelnden Möglichkeit der Datenverschneidung hinzu.

### **Vorgaben von Forschungsorganisationen**

Projektnehmer sind von der Deutschen Forschungsgemeinschaft (DFG) gehalten, die Archivierung von Forschungsdaten für zehn Jahre zu gewährleisten (Grundsätze zur Sicherung guter wissenschaftlicher Praxis, Empfehlung 7: [http://www.dfg.de/aktuelles\\_presse/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_0198.pdf](http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf)), so dass das Ziel des Datenmanagements vor allem die nachhaltige Datenspeicherung und -bereitstellung umfassen muss. Neben den Vorgaben der DFG enthält auch die Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen Vorgaben nationaler und internationaler Forschungsorganisationen (<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>).

### **Bilanz der vergangenen Workshops**

#### **Datenhaltung**

In einer Reihe von Workshops der IT-Fachgruppe konnten Datenbankentwicklungen, wie sie in Deutschland im Bereich Biodiversitäts- und Umweltforschung in den letzten Jahren entstanden, analysiert werden. Dabei zeigte sich, dass, obwohl meist verschiedene Datenbanken eingesetzt werden, die Verwendung von standardisierten Datenaustauschformaten in der Regel nicht ausreichend realisiert ist, was jedoch eigentlich eine Voraussetzung für jegliche IT-Struktur in Projektverbänden sein sollte. Des Weiteren wurde deutlich, dass die Modularisierung der Datenhaltungssysteme nach Datentypen in meist unzureichendem Maße gegeben ist, obwohl eine solche den

flexibleren Umgang auch mit speziellen Datenbeständen ermöglicht. Eine nachhaltige Datensicherung, Dateneingabe und -pflege erwies sich in verteilten Systemen als ebenso möglich wie bei Zentralhaltung. Aufgrund der Möglichkeit der Vergabe abgestufter Zugriffsrechte an verschiedene Autoritäten erschien die dezentrale Datenhaltung in verteilten Systemen für Projektverbünde sogar deutlich geeigneter.

### **Datenauswertung**

Zur Auswertung von in strukturell unterschiedlichen Datenbanken gehaltenen Datenbeständen sollte die Möglichkeit zum Parallelzugriff bestehen. Für die Interoperabilität von Datenbanken (direkt oder indirekt) sind eine saubere Dokumentation und klare Übersichtsstrukturen notwendig, sowie die Verwendung identischer Schlüssel.

Für die verschiedenen Datendomainen (Datentypen) sind bereits konkrete Beispiele vorhanden. Bei der weiteren Verarbeitung der Daten spielen Statistik und deren Visualisierung sowie die Ableitung von Modellen eine zentrale Rolle. Aus diesem Grunde erscheint eine sehr frühzeitige Zusammenarbeit bzw. Absprache mit Modellierungs-Arbeitsgruppen als anstrebenswert, im Zuge derer die Datenhaltestrukturen den Erfordernissen bei Analyse und Modellierung angepasst werden.

### **Kurze Vorstellung der IT-Fachgruppe (Kehl)**

Die IT-Fachgruppe GBIF-Deutschland besteht seit Januar 2004 und ist Teil der Koordinationsstruktur von GBIF-Deutschland. Neben der Veranstaltung von Workshops gehört die Dokumentation von IT-Entwicklungen im Bereich Biodiversitätsinformatik auf technischer Ebene zu den Aufgaben und Zielen. Bisher wurden 10 Workshops durchgeführt an denen nun insgesamt 59 Softwareentwickler und IT-Beauftragte teilgenommen haben. Etwa 50 IT-Entwicklungen konnten in die Dokumentation aufgenommen werden. Protokolle zu den Workshops und online-Formulare für die Dokumentation der IT-Entwicklungen sind auf der Website der IT-Fachgruppe ([http://132.180.63.25/GBIF-ITFG/ITFG\\_welcome.html](http://132.180.63.25/GBIF-ITFG/ITFG_welcome.html)) verfügbar.

### **Beiträge**

Die Folien zu allen Beiträgen sind im PDF-Format auf der Website der IT-Fachgruppe hinterlegt.

1. Bedarfsanalyse bzgl. Daten und Datenhaltung im Rahmen der geplanten Forschungsaktivitäten in den DFG-Exploratorien (Nieschulze)
2. Bereits bestehende Datenhaltungssysteme und IT-Entwicklungen am MPI Jena (Krause)

3. Die Messungen-Datenbank am BayCEER, Status quo und zukünftige Erweiterungen (Archner)
4. Die Komponenten der *Diversity Workbench*: Gesamtkonzept, Modelle, Implementierungen (Weiss)
5. Austauschformate für primäre Biodiversitätsdaten: TDWG Datenstandards und Protokolle (Holetschek)
6. Prozessmanagement für die Integration und Analyse von in heterogenen Datenbanken gespeicherten biologischen Forschungsdaten (Volz)

## **Zentrale Aspekte der Diskussion**

### **Datentypen**

- Die Koordinationsstelle des Projektverbundes sollte bereits zu Projektbeginn bei allen Teilprojekten persönlich die Datenbereiche und Merkmalstypen ihrer geplanten Erfassung direkt abfragen.
- Literaturdaten können in etablierten Literaturverwaltungssystemen eingegeben und verwaltet werden.
- Beschreibenden Daten können zunächst z. B. in Datenhaltesystemen, die das DELTA-Format (<http://delta-intkey.com/>) oder SDD (<http://wiki.tdwg.org/twiki/bin/view/SDD/WebHome>) Austauschformat unterstützen, gesammelt werden.
- Sammlungsdaten können auch in Excel (nach Standardisierung der Datentypen und Vorgabe von „mandatory characters“) gehalten werden und später über Eingabetools in Datenbanken eingegeben werden (z. B. in relationale Sammlungs-Datenbanken wie Specify oder DiversityCollection)

### **Datenerfassung**

- Das Management von Projektverbänden könnte in Erwägung ziehen, mit Vertragspartnern im Rahmen kleiner Projekte die Entwicklung von Standarderfassungstools für spezielle Aspekte zu betreiben (z.B. Beobachtung eines Organismus zu einem bestimmten Zeitpunkt) – entsprechende „hand-held devices“ sind bereits oder demnächst im Handel erhältlich.
- Die Einrichtung einer Helpdesk-Stelle zur Betreuung der Dateneingabe von einzelprojektübergreifenden Kerndaten seitens der Wissenschaftler könnte eine sinnvolle Ergänzung zur Unterstützung der zu schaffenden IT-Strukturen sein.

## Datenhaltung

- Der größte Teil der im Rahmen von Forschungsprojekten anfallenden Daten könnte von bereits bestehenden Entwicklungen abgedeckt werden, so dass sich Entwicklungsarbeiten zur Datenhaltung und -auswertung auf den speziellen Bedarf beschränken kann. Dieser spezielle Bereich muss aber zunächst identifiziert werden.
- Es wäre daher ggf. eine sinnvolle Maßnahme, zunächst eine zentrale Datenbank nur für spezielle Inhalte neu zu entwickeln und für das Gros der zu erwartenden „Standard-Datentypen“ auf bereits bestehende Entwicklungen zu verweisen.
- Die Möglichkeiten von weitgehendem „Outsourcing“ bestimmter Datenbereiche sollten gründlich bedacht werden. Eine Zusammenführung der Daten zu Zwecken der Dokumentation und Auswertung könnte u. a. über BioCASE (<http://www.biocase.org>) und das ABCD Schema (<http://www.bgbm.org/TDWG/CODATA/Schema/>, mit div. Erweiterungen) erfolgen.
- Eine Übersicht über bestehende IT-Entwicklungen (Datenbanken, Schnittstellen, Clients, Wrapper, Portalsoftware etc.) kann die IT-Fachgruppe liefern.
- Nationalpark Berchtesgaden (Helmut Franz, [h.franz@nationalpark-berchtesgaden.de](mailto:h.franz@nationalpark-berchtesgaden.de)) hält die älteste GIS-Datenbank in Deutschland und ist an Kooperationen interessiert.
- Zur datenrechtlichen Seite gibt es Beispiele für Regularien z. B. bei BIOTA AFRICA, die eingesehen werden können (<http://www.biota-africa.org/1024/frames/biota-africa.htm> → Agreements).

## Datenfluss

- Eine Einbeziehung der Prozessmodellierung bzgl. Herkunft, Speicherung, Archivierung, und Darstellung von Primärdaten, deren Prozessierung sowie Darstellung und Archivierung prozessierter Daten wären für einen aus zahlreichen Einzelprojekten bestehenden Verbund von großem Vorteil, v. a. hinsichtlich der möglichen Offenlegung und Vermeidung von Redundanzen. Entsprechende Expertise besitzt der Lehrstuhl Angewandte Informatik IV der Universität Bayreuth. Prof. Jablonski signalisierte Interesse, mit den Koordinatoren des Verbundprojektes der DFG-Exploratorien in einer frühen Phase des Projektes in Kontakt zu treten und ein entsprechendes Vorgehen zu diskutieren.

- Da die Daten aus diesem Projekt gemäß der Berliner Erklärung (<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>) zumindest mittelfristig Wissenschaftlern als auch der Öffentlichkeit zugänglich zu machen sind, ist auf die Nachhaltigkeit der zu etablierenden IT-Strukturen größten Wert zu legen.

### **Nachträgliche Einfügungen, die nicht Gegenstand der Diskussion waren**

- Es bestünde die Möglichkeit die bereits bestehenden Entwicklungen alle zentral in Jena zu installieren. Dabei besteht jedoch ein gewisses Problem darin, dass existierende und potentiell verfügbare Schnittstellen in unterschiedlichen Sprachen programmiert sind. Eine Portierung bestehender Schnittstellen erscheint dabei als eine durchaus sinnvolle Option.
- Die Datenbank aus dem SFB 607 (Projekt C7) kann neben Metadaten auch Primärdaten verwalten, welche auch durch Data-Mining ausgewertet werden können (<http://www.sfb607.de/english/projects/c7/c7.html>)

### ***Angebote zur Kooperation zur Abdeckung bestimmter Datenbereiche***

#### **Kernkonzept Biodiversitätsexploratorien (J. Nieschulze, Dr. S. Pfeiffer)**

Das Informationssystem der Biodiversitätsexploratorien soll die zentrale Erfassung und den Zugang zu den Daten sicherstellen. Zusätzlich soll eine Verknüpfung, Aggregation, statistische Analyse und Visualisierung der Daten bereitgestellt werden. Besondere Bedeutung wird hier der Verknüpfung von Daten über verschiedene räumliche und zeitliche Skalen eingeräumt. Die Visualisierung umfasst sowohl Darstellungen in Karten als auch Methoden der explorativen Datenanalyse.

Die technische Umsetzung sieht einen modularen Systementwurf mit einer XML-basierten, serviceorientierten Architektur (SOA) vor, auf die mittels einer einfachen Web Schnittstelle zugegriffen wird. Bei den Services liegt die Priorität auf dem Einsatz und der Kompatibilität von anerkannten Standards wie GML des OGC und Bio-Case und ABCD.

Für das Zusammenführen und die Analyse von Daten ist deren korrekte Interpretation Grundvoraussetzung. Vor diesem Hintergrund findet zur Zeit eine Metadaten-Schemaentwicklung mit den beteiligten Disziplinen statt. Zum einen sollen mit diesem Schema die Festlegung von Datenformaten und deren inhaltliche Struktur erfolgen. Vorgesehen sind ASCII und eindeutig strukturierte Excel-Dateien. Das Schema soll dabei eine Anzahl von Strukturen (z.B. horizontale oder vertikale Ausrichtung der Variablen) unterstützen und enthält auch Angaben zur Interpretation der Daten. Zum

anderen wird durch das Schema die Kompatibilität mit GML und ABCD sichergestellt, welche als Formate für den Datenzugriff und -austausch vorgesehen sind.

Ein Kernkonzept der Datenverwaltung der Biodiversitätsexploratorien besteht in der Umsetzung der angesprochenen Geschäftslogik in entsprechenden Zugriffs- und Analyseschichten. Dies umfasst neben der Entwicklung und Implementierung von Services die Bereitstellung der dazugehörigen Klienten und deren Einbindung in die SOA. Eine generische Ausrichtung des Systems wird dabei durch Verwendung von individuellen Konfigurationsdateien angestrebt. Die Konzentration der Geschäftslogik in der Mittelschicht wird vor dem Hintergrund der einfacheren Portierung bzgl. des verwendeten Datenbankmanagementsystems (DBMS) gesehen und wird durch die erzielbare Performanz beschränkt. Als DBMS sind solche objektrelationale Produkte im Test, die auch räumliche Datentypen unterstützen.

Bei der Diskussion über Kooperationen zur Abdeckung bestimmter Datenbereiche blieb zum Teil offen, welche Systemvoraussetzungen gegeben sein müssen. Eine Ausnahme ist der PyWrapper, der als Python Anwendung den standardisierten Zugriff via BioCase/ABCD auf eine Vielzahl von Datenbanken realisiert und dessen Einsatz in den Exploratorien in der Testphase ist. Beschränkungen z. B. bzgl. der Erweiterbarkeit oder Modifikation sowie der Leistungsmerkmale möglicher Fremdentwicklungen müssen vor einem Einsatz geklärt werden, da nicht alles im Rahmen des Workshops besprochen wurde. Neben den Primärdaten müssen potentielle Systeme immer auch die projektspezifischen Metadaten gemeinsam verwalten können.

Vor dem Hintergrund des anvisierten SOA werden generell XML basierte Services favorisiert, deren Konfigurationsdateien einem Schema folgen. Voraussetzung eines Einsatzes anderer Systeme ist die Wahrung der Benutzerrechteverwaltung. Wie Zugriffsrechte bei verteilten Hostservern realisiert werden, müsste vorab geklärt werden. Bei verteilter Datenhaltung muss zusätzlich eine einheitliche Versionsverwaltung der Daten durch die eingesetzten Systeme gewährleistet sein.

Die Möglichkeiten und Limitationen der Aggregation und des Zusammenführens von unterschiedlich aufgelösten Daten bei verteilten Datenbanken bedürfen der Klärung. Für die Verschneidung unterschiedlicher Datensätze kann bei verteilten Datenbanken nur bei zumindest zeitweise redundanter Datenhaltung auf die effizienten DBMS-Werkzeuge zurückgegriffen werden. Hierzu müssten die angefragten Daten erst zentral kopiert werden. Engpässe können dabei immer noch durch verfügbare Bandbreiten für den Datendurchsatz entstehen.

Die Diskussion zeigte Entwicklungsbedarf von Anwendungen zur Verwaltung räumlicher Daten in einem DBMS sowie deren Anbindung an OGC konforme Dienste. Hier sollen auf der Datenbankseite die Möglichkeiten, die räumliche Erweiterungen wie z.B. PostGIS oder SpatialExtender bieten, eruiert werden. Für den Zugriff auf solche Datenquellen und deren Einbindung wird der Einsatz von bestehenden und bewährten Entwicklungen wie GeoServer oder Deegree anvisiert.

### **Messungendatenbank (O. Archner)**

Für den Austausch zwischen Wissenschaftlern betreibt das BayCEER eine Datenbank zur Speicherung von Umweltmessreihen.

In Kombination mit dem Klienten „GOAT“ und mehreren ETL (Extraction-, Transformation- and Loading) Modulen deckt sie den gesamten Prozess von der automatischen Datenerfassung im Feld bis hin zur Qualitätssicherung von Messreihen ab.

Die Arbeitsgruppe Datenbanken des BayCEER bietet den Einsatz der Datenbank zur Speicherung von Zeitreihen im Verbundprojekt an. Eine Anpassung der Software an speziellen Anforderungen ist möglich. Bei Bedarf können fehlende Schnittstellen (z. B. BioCASE) durch Mitarbeiter der Arbeitsgruppe realisiert werden.

### **Sammlungs- und Beobachtungsdaten (Dr. D. Triebel)**

Das IT Zentrum an den Staatlichen Naturwissenschaftlichen Sammlungen Bayerns entwickelt generische Software im Bereich Biodiversitätsdaten (siehe z.B. <http://www.snsb.info>, <http://www.diversityworkbench.net>) und vertritt als IT-technische Abteilung zwei Deutsche GBIF-Knoten (Mykologie und Evertebrata II, siehe <http://www.gbif.de>).

Für das Verbundprojekt DFG exploratories könnte speziell die Datenbank-Applikation *DiversityCollection* (sowie assoziierte Applikationen, mit .Net Client) zur Erfassung und Prozessierung von Sammlungs- und Beobachtungsdaten von zentralem Interesse sein und auch bereitgestellt werden. Zusätzliche Bedürfnisse, z. B. hinsichtlich projektspezifischer Import- und Exportfunktionen, könnten in Kooperation mit der IT-Abteilung des Verbundprojektes verwirklicht werden. Sinnvoll wären wohl verteilte Installationen je nach den Anforderungen der Einzelprojekte.

Inwieweit weitere in Entwicklung befindliche Softwarekomponenten der Diversity Workbench (z.B. zur Erfassung beschreibender, ökologischer Daten oder taxonomischer Daten) für die Teilprojekte von Interesse sind, müsste am konkreten Fall analysiert werden.

### **Wrapper (A. Güntsch, J. Holetschek)**

1. Unterstützung bei Installation, Konfiguration und Betrieb der BioCASE Provider Software ("Wrapper")

Wie im Workshop angedeutet wurde, könnte der Austausch der im Projekt erfassten primären Biodiversitätsdaten sowie deren Anbindung an das GBIF-Netzwerk mithilfe der am BGBM entwickelten BioCASE Provider Software (PyWrapper) erfolgen. Die

Daten würden im ABCD-Datenschema (Access to Biological Collection Data) verpackt und mit dem BioCASE-Protokoll abgefragt/versendet werden. Die dafür notwendige Installation und Konfiguration der BPS sowie das Mapping der entsprechenden Datenbanken auf das ABCD-Schema kann durch das BioCASE-Support-Team des BGBM unterstützt werden - per eMail, Telefon oder Remote-Konfiguration und in einzelnen Fällen auch vor Ort.

## 2. Nutzung und Anpassung der am BGBM entwickelten Portal-Softwarepakete

Im Rahmen der BioCASE/SYNTHESYS-Projekte wurden am BGBM zwei Portale zur Abfrage von ABCD-Providern entwickelt, eine sehr einfache Lösung ohne zentrale Datenbank und ein umfangreicheres Interface, das eine eigene Cache-Datenbank nutzt. Beide Portale könnten für den Aufbau von Suchportalen genutzt werden - evtl. mit den entsprechenden Modifikationen.

## 3. Erweiterungen des ABCD-Datenschemas

Sollten die im Projekt anfallenden Daten nicht durch die Elemente des ABCD-Schemas abbildbar sein, so lassen sich diese im "Extensions"-Block verpacken. Bei der dafür notwendigen Definition des Erweiterungsschemas könnte das BGBM seine Erfahrung einbringen.

## 4. Aufbau eines TAPIR-Netzwerkes

Das TAPIR-Protokoll (TDWG Access Protocol for Information Retrieval) ist der gemeinsame Nachfolger der BioCASE- und DiGIR-Protokolle, die derzeit von GBIF unterstützt werden. Das Biodiversitätsexploratorien-Projekt könnte in Zusammenarbeit mit dem BGBM ein TAPIR-Netzwerk aufbauen, das von dessen erweiterten Funktionalitäten profitiert.

## 5. Taxonomischer Backbone.

In Kooperation mit dem BfN und mit europäischen Partnern ist der BGBM für die Einbindung von taxonomischen backbone Daten (Standardlisten der Flora Deutschlands, Euro+Med PlantBase, Fauna Europaea) in die genannten Datenzugangsportale befasst.

## **Prozessanalyse (Prof. Dr. S. Jablonski, B. Volz)**

Die Modellierung und Analyse von Prozessen ist ein anerkanntes und weit verbreitetes Mittel, um komplexe Daten-Integrationsszenarien erfolgreich zu meistern. Die hierbei zugrunde liegenden Szenarien weisen zumeist einen strukturähnlichen Charakter auf: Daten werden – stationär oder auch mobil – erfasst, werden zunächst lokal verwaltet und konsolidiert und müssen anschließend mit Daten anderen Ursprungs integriert werden, d.h. sie werden in einer gemeinsamen Datenbasis konsolidiert und unter Umständen auch miteinander verknüpft. Ein solches Szenario



weist viele Einzelschritte auf, die je nach Anwendungsfall situationsspezifisch konfiguriert werden. Um die Übersicht und Kontrolle über den Gesamtprozess zu erhalten, was insbesondere für die Konsistenzerhaltung der Daten erforderlich ist, werden Prozesse definiert, welche diese Datenintegrationsaufgabe übernehmen.

Der Lehrstuhl für Angewandte Informatik IV der Universität Bayreuth unter Leitung von Prof. Dr.-Ing. Stefan Jablonski bietet daher an, solche Prozessanalysen für Daten-Integrationsszenarien durchzuführen. Die einzelnen Prozesse werden dabei im Rahmen von Interviews mit den betreffenden Institutionen aufgenommen und von Mitarbeitern des Lehrstuhls grafisch modelliert und sukzessive im Rahmen weiterer Interviews verfeinert bis diese den realen Prozess in einer genügend feinen Granularität aufzeigen. Diese Prozesse können dann ausgeführt werden, um den realen Datenaustausch zu steuern. Darüber hinaus bietet der Lehrstuhl praktische Unterstützung bei der Erstellung entsprechender Anwendungen an, die den Datenaustausch und eventuelle Datenkonvertierungen basierend auf Ontologien und Terminologien zum Ziel haben.

Die angebotene Unterstützung ist generell projektbegleitend und teilt sich, wie oben beschrieben, in die zwei Phasen Prozessanalyse und Realisierungsunterstützung/Implementierung ein. Je nach Komplexitäts- und gewünschten Detaillierungsgrad sind erste Ergebnisse der Phase Prozessanalyse nach zwei bis vier Monaten zu erbringen; weitere Prozessanalysen können laufend ergänzt werden. Nach Erstellung erster Prozesse kann deren Umsetzung sogleich beginnen. Eine solche Umsetzung benötigt zwischen 3 und 9 Monaten bei einer mittleren Komplexität der zu integrierenden Datenbestände, der Datentransformationen und der Datentransporte. Üblicherweise werden nachfolgende Umsetzungen schneller zu realisieren sein, da auf bereits implementierte Module zurückgegriffen werden kann. Prozessanalyse und Umsetzung werden in der Regel kontinuierlich bis zum Ende eines Projekts betrieben. Die hier genannten Aufwände beziehen sich aus Erfahrungen aus bereits bearbeiteten Projekten und sind daher nur exemplarisch zu verstehen. Die auszuführenden Tätigkeiten werden typischerweise von einem wissenschaftlichen Mitarbeiter in Zusammenarbeit mit dem Partner aus der Anwendung durchgeführt.

### **Zusammenfassung**

- 1) Es ist zu erwarten, dass im Rahmen von Forschungsaktivitäten des Großprojekts DFG-Biodiversitätsexploratorien unterschiedlichste Daten bzw. Datentypen anfallen werden. Für die Verwaltung, Sicherung und z. T. Auswertung eines Großteils dieser Datenbereiche existiert bereits eine Reihe geeigneter Entwicklungen, die u. U. mit entsprechenden Modifikationen genutzt werden könnten.
- 2) Die Nutzung dieser Entwicklungen könnte auf drei Weisen erfolgen: a) Die entsprechenden Datenbanken und Applikationen werden zentral installiert; b) die Da-

tenbanken werden an verschiedenen Institutionen (z. B. Sammlungen) gehostet und die entsprechenden Daten über Webservices der zentralen Stelle zur Verfügung gestellt; bzw. c) einer Kombination aus a) und b).

3) Eine dezentrale Struktur bei der Datenhaltung steht nicht im Widerspruch mit der Verpflichtung zur langfristigen Vorhaltung und Bereitstellung von Daten. Es bestehen bereits vielfältige technische Möglichkeiten zur Realisierung des Datenflusses.

4) Die Organisation des Datenflusses im Großprojekt Biodiversitätsexploratorien sollte von einer Prozessmodellierung durch Informatiker begleitet werden (siehe Angebot Prozessanalyse).

5) Aus Sicht der IT-Fachgruppe ist es nicht notwendig, innerhalb von zwei Jahren eine endgültige dezentrale Datenstruktur mit Datenaustauschmechanismen etabliert zu haben. Der Aufbau der Strukturen sollte als kontinuierlicher Prozess (unter Beachtung von Rückwärtskompatibilität) verstanden werden. Die IT-technische Vernetzung von Biodiversitätsdaten sollte im nationalen Rahmen stattfinden. Zudem könnten im Rahmen von GBIF-Deutschland aufgebaute Strukturen oder Strukturen, die von den Bundesämtern zur Verfügung gestellt werden, mit einbezogen werden.

6) Es sollten Applikationen zur Dateneingabe für Nutzer, die diese verwenden wollen, zur Verfügung gestellt werden. Auch der Transfer z. B. von Excel-Tabellen in die bestehenden Systeme sollte möglich sein. Für die Betreuung der Nutzer und für die Harmonisierung der Daten ist eine Helpdesk-Stelle sinnvoll.

7) Es sollte bei den einzelnen Projektnehmern rechtzeitig eine Erhebung zur Analyse der benötigten Datentypen erfolgen. Eine Evaluation der bestehenden Datenmodelle wird aus Sicht der IT-Fachgruppe dann wohl zeigen, dass mehr als  $\frac{3}{4}$  der festgelegten Datentypen (teil-)projektübergreifend identisch sind. Die darüber hinausgehenden Daten sollten in den jeweiligen Teilprojekten strukturiert werden. Das Datenmanagement sollte über solche ergänzenden Datentypen informiert sein.

Bayreuth, den 20. September 2007

## **Protokoll zum Workshop XI der IT-Fachgruppe (GBIF-Deutschland):**

### **Wissensbasierten Verwaltung von Daten und der Einsatz von semantischen Systemen im Rahmen der DFG-Exploratorien**

**Beginn:** 17. Dezember 2007 13.30 Uhr  
**Ende:** 18. Dezember 2007 12.00 Uhr  
**Ort:** MPI Biogeochemie Jena

Leitung: Jens Nieschulze, Alexandra Kehl

Protokoll: Alexandra Kehl

Teilnehmer: Oliver Archner (BayCEER, Bayreuth), Calin Arens (MPI, Jena), Dr. Mark Frenzel (Helmholtz-Zentrum für Umweltforschung GmbH – UFZ, Halle), Dennis Heimann (MPI, Jena), Dr. Alexandra Kehl (Uni Bayreuth, GBIF IT-Fachgruppe), Prof. Dr. Birgitta König-Ries (Uni Jena), Prof. Dr. Klaus Küspert (Uni Jena), Prof. Dr. Ralf Kramer (FH Stuttgart), Dr. Felix Mader (Uni Göttingen), Dr. Claus Mückschel (Uni Gießen), Dr. Jens Nieschulze (MPI, Jena), Peter Piehler (PST AG Jena), Herbert Schentz (UBA Wien), Dr. Jens Schumacher (Uni Jena), Cornelia Weist (Uni Gießen).

#### **Programm**

##### Montag, 17. Dezember 2007:

Begrüßung

Vorträge

Gemeinsames Abendessen

##### Dienstag, 18. Dezember 2007:

Weitere Vorträge

Diskussion

## **Begrüßung durch Jens Nieschulze**

### **Kurze Vorstellung der IT-Fachgruppe (Kehl)**

Die IT-Fachgruppe GBIF-Deutschland ist Teil der Koordinationsstruktur von GBIF-Deutschland und besteht seit Januar 2004. ZU den Aufgaben und Zielen gehört neben der Veranstaltung von Workshops die Dokumentation von IT-Entwicklungen im Bereich Biodiversitätsinformatik auf technischer Ebene. Dies ist nun der elfte durchgeführte Workshop, und bisher nahmen insgesamt 59 Softwareentwickler und IT-Beauftragte teil. Etwa 50 IT-Entwicklungen konnten in die Dokumentation aufgenommen werden. Protokolle zu den Workshops und online-Formulare für die Dokumentation der IT-Entwicklungen stehen auf der Website der IT-Fachgruppe ([http://132.180.63.25/GBIF-ITFG/ITFG\\_welcome.html](http://132.180.63.25/GBIF-ITFG/ITFG_welcome.html)) zum Download bereit.

### **Beiträge**

- 11 Jahre zentrales Datenmanagement im Sonderforschungsbereich 299. Ein Anwendungsbeispiel aus der Praxis (Cornelia Weist)
- Metadaten in verteilten Umweltinformations- und Erdbeobachtungssystemen (Ralf Kramer)
- Vom Tabellenblatt zum Informationssystem: Motivation für den Einsatz von MORIS im Rahmen des deutschen Netzwerkes für ökologische Langzeitforschung LTER-D (Mark Frenzel)
- Ontologien und die gemeinsame Verwendung von Daten - Wozu Ontologien.; Ontologien und TDWG (GBIF); Ontologiebasierte Datenhaltung (MORIS) (Herbert Schentz)
- Case Study in Integrating Environmental Databases using Semantic Web Technology (Oliver Archner)
- Teilautomatisierte Datenstrukturanalyse und Software Anforderungen an Applikationsserver (Peter Piehler)
- Bilanz der vorangegangenen Workshops "Datenverwaltung im Rahmen der Forschungsaktivitäten in den DFG-Exploratorien" (Alexandra Kehl)
- Überlegungen zur wissensbasierten Verwaltung und des Einsatz von semantischen Systemen in der Datenverwaltung der DFG-Exploratorien (Dennis Heimann)

Bei der Vorstellung des Projektes Biodiversitäts-Exploratorien wurde vor allem erneut die Heterogenität der verschiedenen beteiligten Teilprojekte und zu erwartenden Da-

ten deutlich. Zur Gewährleistung einer langfristigen Datenhaltung und -bereitstellung, sowie des Datenaustausches zwischen Teilprojekten wurde die Geschäftsordnung um eine „Data sharing policy“ erweitert. Die beteiligten Wissenschaftler müssen ihre erhobenen Daten innerhalb eines Jahres nach Erhebung an das Datenmanagement weiterleiten. Metadaten müssen jedoch jeweils aktuell vor Beginn der Feldarbeit bereitgestellt werden. Die Metadaten sollen neben Verfahrensbeschreibung auch schon die Struktur der geplanten Daten enthalten. 3 Jahre nach Erhebung werden alle Daten veröffentlicht.

Projektnehmer sind außerdem von der Deutschen Forschungsgemeinschaft (DFG) gehalten, die Archivierung von Forschungsdaten für zehn Jahre zu gewährleisten (Grundsätze zur Sicherung guter wissenschaftlicher Praxis, Empfehlung 7: [http://www.dfg.de/aktuelles\\_presse/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_0198.pdf](http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf)), so dass das Ziel des Datenmanagements vor allem die nachhaltige Datenspeicherung und -bereitstellung umfassen muss. Neben den Vorgaben der DFG enthält auch die Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen Vorgaben nationaler und internationaler Forschungsorganisationen (<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>).

Ziel des Datenmanagements in den DFG-Exploratorien ist es, einen effizienten Datenaustausch zwischen den Teilprojekten zu gewährleisten, Möglichkeiten zur Sicherung und Veröffentlichung der Daten zur Verfügung zu stellen, sowie sie an internationale Standards anzubinden. Dabei soll der Zugriff auf die Daten webbasiert möglich sein. Eine Benutzerverwaltung, Service-orientierte Architekturen, ein modulares generisches System und das Zugrundelegen von sogenannten core-Ontologien gehören zu den wichtigsten geplanten Eigenschaften des zukünftigen Datenmanagement-Systems. Noch nicht geklärt ist z. B. inwiefern eine Versionierung der Daten implementiert werden kann.

Als zentrales Problem muss nach wie vor die Gewährleistung der langfristigen Vorkhaltung der Daten gesehen werden. Auch im Rahmen des SFB299, der Ende 2008 nach insgesamt 12 Jahren ausläuft, ist die Frage der langfristigen Datenhaltung weder von Seite der DFG noch der Projektnehmer geklärt. Nach den Erfahrungen im SFB299 muss auch festgestellt werden, dass die Relevanz des Datenmanagements nicht gesehen wird und auch bisher keine Informatiker involviert waren. Es wurden jedoch einige Punkte genannt, wie die Akzeptanz der Nutzer erhöht werden kann.

Vor allem muss eine Motivation für die Nutzung des Systems vorhanden sein. Die Nutzer müssen einen deutlichen Mehrwert erkennen, durch den sie von der Dateneingabe profitieren können. Ein solcher Mehrwert könnte beispielsweise darin bestehen, dass bestimmte Routineauswertungen standardisiert und automatisiert werden können z. B. über R-scripts. Die möglichen Arbeitserleichterungen und Vorteile für die Nutzer können die z. B. in Workshops aber auch in persönlichen Gesprächen

vermittelt werden. Die persönliche Betreuung der Nutzer bei der Dateneingabe sollte eine wichtige Rolle spielen, und die Kommunikation zwischen den Entwicklern und den Nutzern sollte konstant während des gesamten Projektes gewährleistet sein. Für alle erforderlichen Prozesse auf Seiten der Nutzer sollten Beispiele zur Verfügung stehen, an denen man sich orientieren kann. Angebote von Seiten des Datenmanagements an die Nutzer, eine große Benutzerfreundlichkeit und die schnelle Implementierung bestimmter benötigter Funktionalitäten erhöhen die Akzeptanz. Wichtig ist, dass bestimmte Maßnahmen von Beginn des Projektes an greifen.

Die Vorgaben des Datenmanagements der Biodiversitäts-Exploratorien umfassen auch die Eingabe von Metadaten zu allen Arbeiten die in den Exploratorien stattfinden. Erst eine ausreichende Dokumentation ermöglicht den langfristigen Zugriff wie die Interpretierbarkeit der Massendaten. Die Struktur der beschreibenden Metadaten ist maschinenlesbar in XML implementiert. Die ersten Entwicklungen umfassen ein webbasiertes System mit individuell abgestimmten Zugriffsrechten zur Eingabe und zum Zugriff auf Meta- und Massendaten. Die Verwaltung der Meta- und den von ihnen beschriebenen Massendaten erfolgt transparent durch die in den Exploratorien entwickelte Mittelschicht und sichert so Konsistenz. Zugriff auf Massendaten erfolgt ausschließlich über deren Metadaten.

Um effiziente Suchanfragen mit „recall“ und „precision“ Werten jeweils nahe 1 zu erzielen bietet sich nach Erfahrung von Ralf Kramer in z.B. NOKIS an, ein kontrolliertes Vokabular bei den Metadaten einzusetzen. Thesauri erhöhen die Effizienz, sind jedoch sehr zeitaufwendig in Erstellung. Optionale Werte und Freitext in den Metadaten sollten minimiert werden, jedoch sollte es den Benutzern ermöglicht werden, dass eingesetzte Vokabular zu erweitern. Eine Konsistenz- und Redundanzüberprüfung sollte aber immer zentral am besten durch nur eine Person erfolgen, die ebenfalls die Hierarchieeinteilung der Begriffe vornehmen sollte.

Das ontologiebasierte Informationssystem MORIS (Monitoring and Research Information System, <http://www.umweltbundesamt.at/umweltschutz/oeko-system/informationssystem/>) ist gekennzeichnet durch Flexibilität bei der Datenverwaltung. Es besteht die Möglichkeit zur Anpassung der Datenverwaltung in Laufzeit ohne Änderungen des Datenmodells. Die relevanten Metainformationen können zugleich verwaltet werden. Primäre und sekundäre Metainformationen zu den gespeicherten Erhebungen, Objekten und Datenwerten können einfach, effizient und standardisiert abgelegt werden.

Änderungen an Datensätzen werden mitprotokolliert und können somit nachvollziehbar dargestellt und rückgängig gemacht werden. MORIS enthält folgende Bereiche: Objekte, Parameter, Werte, Prozesse und Methoden. Für jeden dieser Bereiche kann der Enduser eigene Klassen definieren. In jedem dieser Bereiche können die Klassen und Instanzen in polyhierarchische Beziehungen zueinander gesetzt werden.

Bei Ontologien handelt es sich nach der Definition von Dr. H. Schentz um Thesauri (controlled vocabularies) mit Relationen zwischen Begriffen (Konzepten), wodurch die Begriffe wechselseitig erklärt werden. Unterschiedliche Daten können nur zusammengeführt und gemeinsam ausgewertet werden, wenn ausreichend Metadaten zur Verfügung stehen. Diese Metadaten müssen in einer gemeinsamen Sprache beschrieben sein. Ontologien in OWL/RDFS und Daten in RDF sind eine gute formale Basis dafür. Ontologien sind erweiterbar, neue Relationen können eingefügt werden, ohne bestehende Strukturen zu zerstören.

Seit Oktober 2006 wird zudem eine Ontologie für TDWG und GBIF entwickelt (<http://wiki.tdwg.org/twiki/bin/view/TAG/TDWGOntology>). Der TAPIR Wrapper kann für eine semantische Vernetzung verwendet werden, doch Portale zur semantischen Vernetzung bestehen noch nicht.

MORIS hat 3 grobe Rollen für die Datenauswahl: Powerselection (Freie Zusammenstellung der Selektionskriterien; für Experten der Oberfläche und des Inhalts), Simple Selection (Vorgefertigte Teilmenge mit vorgefertigten Einschränkungsmöglichkeiten), Durchführen einer fertigen Selection (Jede Auswahl ist speicherbar und kann mit einem Knopfdruck ausgeführt werden).

Wichtig für eine Anwendung im Projekt Biodiversitäts-Exploratorien wäre eine sorgfältige Definition der „core ontology“. Dieser Prozess ist dabei sicherlich der aufwändigste und es müssen dafür alle Disziplinen eingebunden werden. Bei der Entwicklung der „core ontology“ sollte auf die bestehenden Ontologien von TDWQ und MORIS zurückgegriffen werden.

Das System MORIS könnte jederzeit verwendet werden, ist in der jetzigen Version jedoch an Oracle gebunden, soll aber ab Version 2.x weniger abhängig vom DBMS sein. Die Entwickler orientieren sich mehr in Richtung PostgreSQL.

Betont wurde noch einmal, dass nur eine direkte Verknüpfung der Metadaten mit den Daten eine Automatisierung von Prozessen möglich macht. Aggregationen von Daten, Datentypkonvertierungen und Datenvergleiche sind nur möglich wenn wirklich detaillierte Metadaten erhoben wurden. Der Mehrwert dieser zunächst aus Sicht der Nutzer zusätzlichen Arbeit muss klar erkennbar sein. Eine Festlegung auf bestimmte Formate ist dabei unumgänglich.

Gemeinsame Semantiken für verschiedene Disziplinen sind von großer Bedeutung, denn Definitionen können wieder verwendet werden (z.B. bei der Laborarbeit). Methoden werden normalerweise länger beibehalten. Hier ist aber auch zu bedenken, dass in manchen Disziplinen (z.B. Vegetationskunde) die angewandte Methode durchaus vom Anwender abhängig sein kann. Eine Einigung auf wenige gemeinsam verwendete Methoden wird im Projekt Biodiversitäts-Exploratorien als nicht realistisch angesehen.

## **Fazit**

- Technische Probleme sind mehr oder wenig einfach zu lösen. Die größte Herausforderung liegt in der inhaltlichen Bestimmung der Geschäftslogik.
- Der Einsatz von Ontologien im Austausch und der Integration von Daten ist ein aktuelles Forschungsthema und en vogue, auch wenn einige Ansätze recht alt und z.B. mit CORBA schon früher abgedeckt wurden. Nicht überall wo Ontologie drauf steht ist auch Ontologie drin.
- Die Heterogenität der in den Biodiverdiversitäts-Exploratorien beteiligten Projekte wird sich in den produzierten Daten widerspiegeln. Die beschreibenden Metadaten werden durchgängig nach XML-Schemas standardisiert und transparent mit den Massendaten verknüpft. Der Aufbau eines kontrollierten Vokabulars erscheint sinnvoll.
- Hauptaufgabe des Datenmanagements in den Exploratorien ist die Verwaltung, Bereitstellung und Visualisierung der erhobenen Daten mit Hinblick auf Austausch zwischen den Projekten. Diese Vorgabe bedingt zusätzlich die Integration und teilweise die Analyse und Auswertung von Daten. Hier liegt der Mehrwert des Gesamtprojektes als auch für die individuellen Benutzer. In beiden Fällen müssen Daten und Analysemethoden integriert werden. Die in den Projekten verwendeten Begrifflichkeiten sind jedoch erfahrungsgemäß Disziplin spezifisch. Erst eine Abstrahierung und Formalisierung der Begrifflichkeiten erlaubt deren automatisierte Verwendung in der Datenintegration und -auswertung. Hier setzen die von MORIS oder TDWG eingesetzten Ontologien an und machen sie für die Exploratorien interessant. Der Einsatz beider Ansätze ebenso wie die Verwendung von MORIS wird geprüft.
- Die langfristige Haltung, Pflege und Bereitstellung von Projektdaten über den Förderungszeitraum wird i.d.R. von den Projektträgern nicht berücksichtigt.

Bayreuth, den 20. Dezember 2007



## Anlage 2: Gesamtliste der Workshopteilnehmer

<b>Name</b>	<b>Telefon</b>	<b>E-Mail</b>
Andreas Allspach		andreas.allspach@senckenberg.de
Oliver Archner	0921/555727	oliver.archner@bayceer.uni-bayreuth.de
Calin Arens	03641/576137	carens@bgc-jena.mpg.de
Anna Augustin	040/42816262	anna.augustin@botanik.uni-hamburg.de
Andreas Beck	089/17861266	beck@bsm.mwn.de
Evelin Boos		evelin.boos@snafu.de
Minnattallah Boutros	0931/8884354	boutros@biozentrum.uni-wuerzburg.de
Stefan Dressler	069/970751	Stefan.Dressler@senckenberg.de
Jörg Ewald	08161/715909	joerg.ewald@fh-weihenstephan.de
Manfred Finckh	040/42816260	mfinckh@botanik.uni-hamburg.de
Frauke Fischer	0931/8884365	fischer@biozentrum.uni-wuerzburg.de
Helmut Franz		h.franz@nationalpark-berchtesgaden.de
Mark Frenzel	0345/5585304	mark.frenzel@ufz.de
Alexander Gröngroeft	040/428384395	a.groengroeft@ifb.uni-hamburg.de
Anton Güntsch	030/83850166	a.guentsch@bgbm.org
Gregor Hagedorn	030/83042220	G.Hagedorn@bba.de
Andrea Hahn	+45 35321491	ahahn@gbif.org
Dennis Heimann	03641/576108	dheimann@bgc-jena.mpg.de
Andreas Heinecke	0209/9596788	amh@hi-soft.de
Andreas Hirner	08153/283389	andreas.hirner@dlr.de
Jörg Holetschek	030/83850150	j.holetschek@bgbm.org
Joachim Holstein	0711/8936234	holstein.smns@naturkundemuseum-bw.de
Stefan Holzheu	0921/555720	stefan.holzheu@bayceer.uni-bayreuth.de
Jürgen Hoppe	0731/5022703	juergen.hoppe@uni-ulm.de
Jan Ingenhaag	089/17861303	jan.ingenhaag@lrz.uni-muenchen.de
Alexandra Kehl	0921/552455	alexandra.kehl@uni-bayreuth.de
Manfred Keil	08153/281377	manfred.keil@dlr.de
Andreas Kohlbecker	030/83042261	kohlbecker@web.de
Birgitta König-Ries		koenig@informatik.uni-jena.de
Ralf Kramer		ralf.kramer@hft-stuttgart.de
Reiner Krause	03641/576318	rkrause@bgc-jena.mpg.de
Alexander Kroupa	0921/98756	alexander.kroupa@s2you.de
Klaus Küspert		kuespert@informatik.uni-jena.de

Asif Rashid Lakehsar	089/17861304	lakehsar@bsm.mwn.de
Karl-Heinz Lampe	0228/9122290	k.lampe.zfmk@uni-bonn.de
Ekaterina Langer	030/83875234	atanasso@inf.fu-berlin.de
Thorsten Ludwig	0731/5022699	thorsten.ludwig@biologie.uni-ulm.de
Felix Mader		fmader@ufobi6.uni-forst.gwdg.de
Rudolf May	0228/8491-1441	rudolf.may@bfm.de
Lothar Menner	069/7542363	Lothar.Menner@senckenberg.de
Gerhard Muche	040/42816395	gerhard.muche@botanik.uni-hamburg.de
Claus Mücke		Claus.Mueckschel@agrار.uni-giessen.de
Andrejs Naumovs	0531/2616327	andrejs.naumovs@dsmz.de
Dieter Neubacher	089/72059631	dneu@arcor.de
Jens Nieschulze	03641/576138	jniesch@bgc-jena.mpg.de
Clemens Oertel		oertel@uni-tuebingen.de
Johannes Penner	09318884378	penner@biozentrum.uni-wuerzburg.de
Simone Pfeiffer	0331/9771905	simone.pfeiffer@uni-potsdam.de
Peter Piehler		peter.piebler@pst.ag
Gerhard Rambold	0921/552453	gerhard.rambold@uni-bayreuth.de
Dominik Röpert	030/83850172	d.roepert@bgbm.org
Herbert Schentz		herbert.schentz@umweltbundesamt.at
Klaus Schild	030/83875228	schild@inf.fu-berlin.de
Marco Schmidt	069/97075615	marco.schmidt@senckenberg.de
Andreas Schnock	030/20938558	andreas.schnock@museum.hu-berlin.de
Christoph Schultz	0931/8884797	christoph.schultz@dlr.de
Stephanie Sobek		
Henning Sommer		
Wiltrud Spiesberger		
Dirk Striebing	030/20938851	dirk.striebling@museum.hu-berlin.de
Peter Strobl		pstrobl@cosy.sbg.ac.at
Lutz Suhrbier	030/83875227	suhrbier@inf.fu-berlin.de
Dagmar Triebel	089/17861252	triebel@bsm.mwn.de
Melanie Vogel	0931/8884797	melanie.vogel@mail.uni-wuerzburg.de
Bernhard Volz	0921/557343	bernhard.volz@uni-bayreuth.de
Markus Weiss	089/17861303	weiss@bsm.mwn.de
Cornelia Weist		Cornelia.Weist@agrار.uni-giessen.de
Michael Wiedemann	0731/5022699	michael.wiedemann@uni-ulm.de

## Anlage 3: Gesamtliste der Workshop-Beiträge

Die Beiträge sind als pdf-Dateien auf der Homepage der IT-Fachgruppe hinterlegt.

<b>Teilnehmer</b>	<b>Titel</b>	<b>WS</b>
Anna Augustin (BIOTA, Hamburg)	Anbindung des Herbarium Hamburgense an GBIF als Ausgangslösung einer online BIOTA Datenbasis	IV
Oliver Archner (BayCEER, Bayreuth)	Die Messungen-Datenbank am BayCEER: Status quo und zukünftige Erweiterungen	X
	Case study in integrating environmental databases using semantic web technology	XI
Stefan Dressler (BIOTA W11, Frankfurt)	Westafrika Sammlungsdaten im BIOTA Projekt W11	IV
Mark Frenzel (UFZ, Halle):	Vom Tabellenblatt zum Informationssystem: Motivation für den Einsatz von MORIS im Rahmen des deutschen Netzwerkes für ökologische Langzeitforschung LTER-D	XI
Anton Güntsch (BGBM, Berlin)	Datenaustauschformate für Sammlungsdaten: Das ABCD-Schema	II
Gregor Hagedorn (BBA, Berlin)	TDWG 2004 - Fortschritte bei Datenaustausch-Standards	II
Andrea Hahn (GBIF, Berlin)	Daten und Datenflüsse in GBIF Botanik	II
	BioCASE und ABCD	IV
Dennis Heimann (MPI, Jena)	Überlegungen zur wissensbasierten Verwaltung und des Einsatzes von semantischen Systemen in der Datenverwaltung der DFG-Exploratorien	XI
Andreas Heinecke (Fachhochschule Gelsenkirchen)	Modellierung von BIOTA-Daten nach CIDOC CRM	VI
Jörg Holetschek (GBIF, Berlin)	GBIF Architekturgrundlagen	IV
	Austauschformate für primäre Biodiversitätsdaten: TDWG Datenstandards und Protokolle	X
Jan Ingenhaag (GBIF Mykologie, München)	Wrapper Installation am GBIF Knoten Mykologie	IV
Alexandra Kehl (BIOTA S04, Bayreuth)	Datenbanken und Datenbankclient DiversityNavigator Biota Süd	IV
Alexandra Kehl (Uni Bayreuth, GBIF IT-Fachgruppe)	Kurze Vorstellung der IT-Fachgruppe	X, XI
Ralf Kramer (FH Stuttgart)	Metadaten in verteilten Umweltinformations- und Erdbeobachtungssystemen	X
Reiner Krause (MPI, Jena)	Bereits bestehende Datenhaltungssysteme und IT-Entwicklungen am MPI Jena	X

Karl-Heinz Lampe (BIOTA Ost, Bonn)	CIDOC CRM Mapping	IV
Ekaterina Langer (Institut für Informatik, Freie Universität Berlin)	XML-Sicherheitsmechanismen für den Datentransfer in GBIF	II
Thorsten Ludwig (SysTax, Ulm)	SysTax – Datenfluss	II
	SysTax	IV
Gerhard Muche (BIOTA, Hamburg)	Datentransfer als Scharnier zwischen <i>datamining</i> und <i>data consolidation</i>	II
	Anspruch der Biodiversitätsforschung an Daten als Argumentationsgrundlage	IV
	BIOTA Datenhaltung	V
	Meilensteine zum Ablauf BIOTA-GBIF-Datenverbund	VI
Jens Nieschulze (MPI, Jena)	Bedarfsanalyse bzgl. Daten und Datenhaltung im Rahmen der geplanten Forschungsaktivitäten in den DFG Biodiversitäts-Exploratorien	X
	Datenmanagement in den DFG Biodiversitäts-Exploratorien – aktueller Stand und Planung	XI
Clemens Oertel (Uni Tübingen)	MoDaB Tübingen & Einführung in XPath/XQuery	II
Herbert Schentz (UBA, Wien)	Ontologien und die gemeinsame Verwendung von Daten	XI
Klaus Schild (Institut für Informatik, Freie Universität Berlin)	Webservices: Eine kritische Einführung	II
	XSLT: Transformation von XML-Dokumenten	
Marco Schmidt (BIOTA W11, Frankfurt)	VegDa – eine Datenbank für Vegetationsaufnahmen und botanische Inventare	IV
	Datenvielfalt in der Vegetationsdatenbank VegDa	VI
Christoph Schultz (BIOTA W01 und S01, Würzburg)	Datenstrukturen in den BIOTA Teilprojekten W01 und S01	IV
Peter Strobl (Uni Salzburg)	Austauschstandard für Vegetationsdatenbanken ESveg	VI
Berhard Volz (Uni Bayreuth, Angewandte Informatik IV)	Prozessmanagement für die Integration und Analyse von in heterogenen Datenbanken gespeicherten biologischen Forschungsdaten	X
Markus Weiss (IT-Zentrum der SNSB, München)	Vorstellung DiversityWorkbench	V
	Datenbankmodell DiversityEcology	VI
	Die Komponenten der DiversityWorkbench: Gesamtkonzept, Modelle, Implementierungen	X

## Anlage 4: Merkmalslisten der HTML-Formulare

### Merkmalsliste des HTML-Formulars „IT-Entwicklungen“

Erläuterungen:

TE: Text, UM: unordered multistate, OM: ordered multistate

	Merkmalstyp
<u>Eingabedatum für dieses Formular</u>	
1. Datum der Ersteingabe (z.B. 03.09.2004):	TE
2. Revisionsdatum (z.B. 03.09.2004; 14.10.2004):	TE
<u>Allgemeines</u>	
3. Name der Institution:	TE
4. Verantwortlicher für IT-Entwicklung (Name und Email):	TE
5. Beteiligte Softwareentwickler (Namen und Email):	TE
<u>Verfügbarkeit und Lizenzen</u>	
6. Ist Ihre IT-Entwicklung bzw. Ihr IT-Produkt für Dritte verfügbar?	UM
1. Ja	
2. Nein	
7. Lizenztyp:	UM
1. OSI-zertifizierte open source Software	
2. Andere Lizenz	
3. Lizenzpolitik ungeklärt	
4. Produkt für Dritte nicht verfügbar	
8. Bezeichnung der Lizenz (z.B. GPL, LGPL):	TE
9. Copyright für Eigenentwicklungen am Softwareprodukt liegt bei:	TE
<u>Eckdaten zum IT-Produkt</u>	
10. Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten:	UM
1. Wissenschaftliche Namen	
2. Trivialnamen	
3. Synonyme	
4. Sammlungsbelege (Naturgeschichtliche Sammlungen)	
5. Beobachtungsdaten (Monitoring, Erfassungen)	
6. Bilddaten	
7. Audio-Daten	
8. Geographische Daten	
9. Geologische Daten	
10. Meteorologische Daten	
11. Molekulare Daten	
12. Populationsdaten	
13. Ökologische Daten	
14. Schutzstatus	
15. Beschreibende Daten	
16. Daten zur Präparation und Aufbewahrung	
17. Literaturverwaltung	
18. Linksammlung	
19. Andere	
11. Nutzung des IT-Produkts:	UM
1. Über Client-Server-Anbindung	
2. Lokal	
3. Ausschließlich über www	
12. Art der IT-Entwicklung:	UM
1. Datenbankanwendung	
2. Client-Software	
3. Anwenderprogramm	
4. Web-Schnittstelle	
5. Webservice, Wrapper	
6. Andere	
13. Kurze Beschreibung:	TE
14. Zugangsseite (URL):	TE
15. Besteht das IT-Produkt aus mehreren unabhängigen Modulen?	UM
1. Ja	
2. Nein	
16. Für IT-Produkt verwendbare HTTP Server:	UM
1. Apache	
2. Microsoft IIS	
3. NCSA HTTPd	
4. Andere	
17. Entwicklungsstand des IT-Produkts (aktuelle Versionsnummer)	UM

angeben):

1. In Planung
  2. In Entwicklung
  3. Beta Testversion
  4. Stabile Version
18. In der Benutzeroberfläche verwendete Sprache: UM
1. Nur deutsch
  2. Nur englisch
  3. Mehrsprachig (bitte Sprachen angeben)
  4. Andere
19. Ist ein Hilfesystem oder eine Benutzeranleitung verfügbar? UM
1. Ja, online
  2. Ja, offline
  3. In Bearbeitung
  4. Nein
20. Ist eine technische Dokumentation vorhanden bzw. verfügbar? UM
1. Ja, online
  2. Ja, offline
  3. In Bearbeitung
  4. Nein

#### Programmiersprache und Quelltext

21. Für das IT-Produkt verwendete Sprache(n): UM
1. C/C++
  2. Java
  3. VB
  4. .Net
  5. Eiffel
  6. Python
  7. Andere
22. Entwicklungsumgebung(en): TE
23. Sind die Quelltexte der IT-Entwicklung für Dritte verfügbar? UM
1. Ja
  2. Nein
24. Sind die Quelltexte kommentiert? UM
1. Ja
  2. Nein
25. In welcher Sprache sind die Quelltexte kommentiert? UM
1. Deutsch
  2. Englisch
  3. Andere
26. Quelltext (URL): TE
27. Quelltext (CVS Account): TE

#### Datenbankentwicklungen

28. Verwendetes Datenbank-Verwaltungssystem (Version angeben): UM
1. Oracle
  2. MySQL
  3. MiniSQL
  4. PostgreSQL
  5. MS SQL Server
  6. MS Access
  7. MS Works
  8. Filemaker
  9. Andere
29. Gibt es ein publiziertes Datenmodell? UM
1. Ja
  2. Nein
  3. In Bearbeitung
30. Bezeichnung des Datenmodells: TE
31. Datenbank Import-Formate: UM
1. XML
  2. Feldbegrenzer-limitierter Text
  3. Andere

32. Datenbank Export-Formate: UM
1. XML
  2. Feldbegrenzer-limitierter Text
  3. Andere
33. Datenbank-Datenbank-Content-Transfer über: UM
1. Import/Export-Dateien
  2. ODBC
  3. JDBC
  4. Andere
34. Integration von "authority files" (z.B. Gazetteers)? UM
1. Ja
  2. Nein
35. Welche "authority files" können verwendet werden? TE
- Datenbanken und Web
36. Ist eine Datenbankabfrage über das WWW möglich? UM
1. Ja, mit statischer Dokumentdarstellung
  2. Ja, mit dynamischer Dokumentengenerierung
  3. In Bearbeitung
  4. Nein
37. Ist die Datenbankabfrage über das WWW eingeschränkt? UM
1. Ja
  2. Nein
38. Ist es möglich die Datenbankinhalte über das WWW zu editieren? UM
1. Ja, uneingeschränkt
  2. Ja, von bestimmten Nutzern/Bearbeitern
  3. Ja, bestimmte Inhalte
  4. In Bearbeitung
  5. Nein
39. Zugriff auf die Datenbank über (Schnittstelle zur Datenbank): UM
1. ODBC
  2. JDBC
  3. PL/SQL (Oracle procedural language)
  4. .Net
  5. Andere
40. Realisierung (Technologie) der Web Schnittstelle: UM
1. CGI
  2. API
  3. SSI
  4. ASP
  5. Java servlets
  6. Java applets
  7. JSP
  8. Portlets
  9. .Net
  10. Andere
41. Verwendete Programmiersprache für Datenbank-Web Schnittstelle: UM
1. PHP
  2. Perl
  3. ColdFusion
  4. Python/Zope
  5. JavaScript
  6. VBScript
  7. JScript
  8. Java
  9. C/C++
  10. Tcl
  11. .Net
  12. Andere
- Clients und Anwendungsprogramme
42. Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? (Version angeben) TE

43. Welche Laufzeitumgebung ist für das IT-Produkt notwendig?  
(Version angeben) TE
- Webservice
44. Verwaltung des Web Service über: UM  
 1. UDDI  
 2. WS-Inspection  
 3. Andere
45. Verbindungsprotokoll für Webservices: UM  
 1. SOAP/RPC  
 2. Andere
46. Interface für Wrapper: UM  
 1. DiGIR  
 2. BioCASE
- Sonstiges
47. Bisherige externe Nutzer Ihrer IT-Entwicklung (Personen, Institutionen): TE
48. Geplante Erweiterungen für Ihre Entwicklung: TE
49. Sie würden gerne an einem (oder mehreren) Workshops zu folgenden Themen teilnehmen: UM  
 1. Spezifikationen für den Datenaustausch  
 2. Web-Services  
 3. Datentransfer und Sicherheitskonzepte  
 4. Spezifische Probleme der Interoperabilität verschiedener Systeme (z.B. GIS Anbindung)  
 5. Datenbanken für verschiedene Inhalte (z.B. beschreibende Daten)  
 6. GBIF-D und GBIF international: Integration und Interoperabilität
50. Weitere gewünschte Themen für einen Workshop: TE
51. Weitere Fragen, die in dieses Formular aufgenommen werden sollten, sowie Korrekturvorschläge: TE

## **Merkmalsliste des HTML-Formulars „Institution“**

Erläuterungen:

TE: Text, UM: unordered multistate, OM: ordered multistate, IN: ganze Zahl

### Eingabedatum für dieses Formular

1. Datum (z.B. 04.09.2004): TE
2. Revisionsdatum (z.B. 04.09.2004; 20.09.2004): TE

### Allgemeines zur Institution

3. Art der Institution: UM  
 1. Universität oder Hochschule  
 2. Naturgeschichtliche Sammlung oder Museum  
 3. Forschungsinstitut  
 4. Unternehmen  
 5. Privatperson(en)
4. Adresse: TE
5. Homepage (URL): TE
6. Abteilung: TE

### Projekte

7. Laufende Projekte im Bereich Biodiversitätsinformatik mit eigener IT-Entwicklung: TE
8. Laufende Projekte im Bereich Biodiversitätsinformatik ohne eigene IT-Entwicklung: TE
9. Bereits beendete Projekte: TE
10. Informationen über IT-Projekte (URL): TE

### IT-Entwicklungen

11. IT-Koordinator(en) (Name(n)und Email): TE
12. Anzahl interner Softwareentwickler: IN



13. Anzahl externer Softwareentwickler: IN
- GBIF-Deutschland
14. Einbindung in GBIF-Struktur? UM
1. Ja
  2. Nein
15. Sollen Ihre Daten in Zukunft für einen der GBIF-Deutschland Knoten verfügbar gemacht werden? UM
1. Nein
  2. Ja, für Viren und Prokaryonten
  3. Ja, für Pflanzen (+ Protisten)
  4. Ja, für Pilze
  5. Ja, für Evertabrata I (Insecta)
  6. Ja, für Evertabrata II
  7. Ja, für Evertabrata III
  8. Ja, für Vertebrata
- Daten
16. Welche Typen von Biodiversitätsdaten werden derzeit erhoben (digital)? UM
1. Wissenschaftliche Namen
  2. Trivialnamen
  3. Sammlungsbelege
  4. Beobachtungsdaten (Monitoring, Erfassungen)
  5. Bilddaten
  6. Geographische Daten
  7. Geologische Daten
  8. Meteorologische Daten
  9. Molekulare Daten
  10. Populationsdaten
  11. Ökologische Daten
  12. Schutzstatus
  13. Beschreibende morphologische und anatomische Daten
  14. Daten zur Präparation und Aufbewahrung
  15. Literaturverwaltung
  16. Linksammlung
  17. Andere
17. Welche Typen von Biodiversitätsdaten sollen in näherer Zukunft erhoben werden (digital)? UM
1. Wissenschaftliche Namen
  2. Trivialnamen
  3. Sammlungsbelege
  4. Beobachtungsdaten
  5. Bilddaten
  6. Geographische Daten
  7. Geologische Daten
  8. Meteorologische Daten
  9. Molekulare Daten
  10. Populationsdaten
  11. Ökologische Daten
  12. Schutzstatus
  13. Beschreibende morphologische und anatomische Daten
  14. Daten zur Präparation und Aufbewahrung
  15. Literaturverwaltung
  16. Linksammlung
  17. Andere
18. Art der Datenspeicherung: UM
1. Relationale Datenbank (z.B. MS Access)
  2. Strukturiert, Tabellenkalkulation (z.B. Excel)
  3. Unstrukturiert, als Textdateien
19. Ort der Datenspeicherung: UM
1. Am eigenen Institut
  2. An anderer Stelle (Hosting)
  3. An anderer Stelle (Hosting), aber es ist langfristig geplant, eigene IT-Strukturen zu entwickeln

20. Ort der Datenpräsentation bzw. Speicherort der Webseiten: UM
1. Am eigenen Institut
  2. An anderer Stelle (Hosting)
  3. An anderer Stelle (Hosting), aber es ist langfristig geplant, eigene IT-Strukturen zu entwickeln
21. Externe Kooperationen bezüglich der Datenverwaltung mit TE  
(Personen, Institutionen):
- Technische Strukturen
22. Verwendetes Betriebssystem auf dem/den Arbeitsplatzrechner(n) UM
1. MS Windows
  2. Linux
  3. Unix
  4. Sun/Solaris
  5. Apple Macintosh
  6. MAC OS-X
  7. Andere
23. Ist eine Migration zu einem anderen Betriebssystem geplant? UM
1. Nein
  2. Ja, nach MS Windows
  3. Ja, nach Linux
  4. Ja, nach Unix
  5. Ja, nach Sun/Solaris
  6. Ja, nach Apple Macintosh
  7. Ja, nach MAC OS-X
24. Anzahl der Datenbankserver: IN
25. Anzahl der für Biodiversitätsdaten verwendeter HTTP-Server: IN
26. Verwendeter HTTP-Server: UM
1. Apache
  2. Microsoft IIS
  3. NCSA HTTPd
  4. Andere
27. Verwendetes Betriebssystem auf Ihrem Datenbank- bzw. HTTP-Server UM
1. MS Windows
  2. Linux
  3. Unix
  4. Sun/Solaris
  5. Apple Macintosh
  6. MAC OS-X
  7. Andere
28. Geschwindigkeit der Internetanbindung: UM
1. 56/64 KBit/s
  2. 256 KBit/s
  3. 512 KBit/s
  4. 1 MBit/s
  5. 2 MBit/s
  6. 10 MBit/s
  7. 100 MBit/s
  8. 1 GBit/s
  9. Andere
29. Sind bereits Wrapper installiert? UM
1. Ja
  2. Nein
30. Welche Wrapper wurden oder werden installiert? UM
1. PyWrapper (ABCD-Schema)
  2. Andere (bitte nennen)
- Sonstiges
31. Was müsste ein Datenbankprogramm für Ihre Biodiversitätsdaten können? (Wünsche an "optimale" Software) TE
32. Sie würden gerne an einem (oder mehreren) Workshops zu folgenden Themen teilnehmen: UM
1. Spezifikationen für den Datenaustausch
  2. Web-Services (Wrapper, RPCs)
  3. Datentransfer-Sicherheitskonzepte

4. Spezifische Probleme der Interoperabilität verschiedener Systeme (u. a. GIS Anbindung)
5. Datenbanken für verschiedene Inhalte (z.B. deskriptive Daten)
6. GBIF-D und GBIF-international: Integration, Interoperabilität
33. Weitere gewünschte Themen für einen Workshop:
34. Weitere Fragen, die in dieses Formular aufgenommen werden sollten, sowie Korrekturvorschläge:

TE  
TE

## Anlage 5: Auflistung der dokumentierten IT-Entwicklungen

	<b>Bezeichnung der IT-Entwicklung</b>	<b>Institution(en)</b>	<b>Stand</b>
1	AlgaTerra Datenbank + Webinterface	Botanischer Garten und Botanisches Museum Berlin-Dahlem	Januar 2008
2	BIG (Bundesinformationssystem Genetische Ressourcen): Teilprojekt Wildpflanzen	Bundesamt für Naturschutz im Verbund mit anderen Institutionen (RUB, IPK, ZADI)	April 2007
3	Biodat	Museum Alexander Koenig/Bonn, Museum für Naturkunde/Berlin	Juli 2007
4	BIOTABase-Family	Universität Hamburg Biozentrum Klein Flottbek & Botanischer Garten	Februar 2007
5	Bogart	Botanischer Garten und Botanisches Museum Berlin-Dahlem	Juni 2006
6	CeDaMar-Webinterface	Forschungsinstitut Senckenberg (FIS), Abteilung Deutsches Zentrum für Marine Biodiversitätsforschung (DZMB)	September 2004
7	DALI - Database of Ascomycete Literature	Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT-Zentrum	Juli 2007
8	Datenbank Herbarium Hamburgense	Universität Hamburg Biozentrum Klein Flottbek & Botanischer Garten	Februar 2007
9	Digitale Belege	Botanischer Garten und Botanisches Museum Berlin-Dahlem	September 2004
10	DiversityCollection	Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT-Zentrum	Juli 2007
11	DiversityExsiccatae	Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT-Zentrum	Juli 2007
12	DiversityNavigator	Universität Bayreuth, Lehrstuhl Pflanzen-systematik, Abt. Mykologie	Februar 2007
13	DiversityTaxonNames	Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT-Zentrum	Juli 2007
14	DiversityWorkbench	Biologische Bundesanstalt für Land- und Forstwirtschaft	Juni 2006
15	EmailDirectory	Staatliche Naturwissenschaftliche Samm-	Juli 2007

		lungen Bayerns, IT-Zentrum	
16	FloraWeb	Bundesamt für Naturschutz	April 2007
17	Foto Gallery Internet	Universität Hamburg Biozentrum Klein Flottbek & Botanischer Garten	September 2004
18	Genbankinformationssystem (GBIS)	Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) Gatersleben	Juli 2007
19	GLOBIS-D Webpräsentation	Staatliches Museum für Naturkunde Stuttgart	April 2007
20	HerbarDatenbank des IPK Gatersleben	Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) Gatersleben	Juli 2007
21	interne MS Access Entwicklung	Universität Bonn	September 2004
22	Internet Datenbank Biota-W	Universität Bonn	September 2004
23	JsLit (Version 8, vormals 6)	Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT-Zentrum	Juli 2007
24	LinkCollection	Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT-Zentrum	Juli 2007
25	Mansfeld's World Database of Agricultural and Horticultural Crops (Mansfeld-Datenbank)	Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) Gatersleben	Juli 2007
26	Materialarchiv Forschungsschiff Meteor	Forschungsinstitut Senckenberg (FIS), Abteilung Deutsches Zentrum für Marine Biodiversitätsforschung (DZMB)	September 2004
27	MS Access Anwendung für Prokaryota Knoten	DSMZ - Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH	September 2004
28	MyrChar (Version 1)	Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT-Zentrum	Juli 2007
29	MyrKar (Version 6)	Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT-Zentrum	Juli 2007
30	NaviKey 4 – a Java applet and application for accessing descriptive data coded in DELTA format	Universität Bayreuth, Lehrstuhl Pflanzen-systematik, Abt. Mykologie	Juli 2007
31	NeoFlora-Converter	Bundesamt für Naturschutz	Oktober 2004
32	Portal Virtuelles Herbarium (Botanischer Knoten)	Botanischer Garten und Botanisches Museum Berlin-Dahlem	Juni 2006
33	PyWrapper	Botanischer Garten und Botanisches Museum Berlin-Dahlem, Abt. Biodiversitätsin-formtik	Juni 2006
34	science4you online Monitoring	science + communication, Norbert Hirneisen	April 2007

35	SeSam (Senckenbergisches Sammlungsverwaltungssystem)	Forschungsinstitut und Naturmuseum Senckenberg, Frankfurt/Main	April 2004
36	Specify	Specify Software Project, University of Kansas Biodiversity Research Center (Betreuung in DE: BGBM Berlin-Dahlem)	Juli 2006
37	SPICE-Wrapper	Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT-Zentrum	Juli 2006
38	SYNTHESYS Data Quality Tool	Botanischer Garten und Botanisches Museum Berlin-Dahlem, Abt. Biodiversitätsinformatik	Juli 2007
39	SysMyr (Version 4)	Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT-Zentrum	Juli 2007
40	SysTax	Institut für Systematische Botanik und Ökologie, Universität Ulm	Mai 2007
41	TaxDB	Forschungsinstitut Senckenberg (FIS), Abteilung Deutsches Zentrum für Marine Biodiversitätsforschung (DZMB)	September 2004
42	Taxonomischer Webeditor für Berlin Modell Datenbanken	Botanischer Garten und Botanisches Museum Berlin-Dahlem, Abt. Biodiversitätsinformatik	Juni 2006
43	Unitloader	Botanischer Garten und Botanisches Museum Berlin-Dahlem, Abt. Biodiversitätsinformatik	Juni 2006
44	VegDa	Forschungsinstitut Senckenberg, J.W.Goethe-Universität Frankfurt	Juli 2007
45	Visual Plants	Universität Hohenheim, Institut für Botanik	Juli 2007
46	WISIA	Bundesamt für Naturschutz	April 2007
47	Zoologische und botanische Funddaten	Nationalpark Berchtesgaden	April 2007

## **Anlage 6: Beschreibung der dokumentierten IT-Entwicklungen**

Alphabetische Liste, maschinell erstellt.

### **1 AlgaTerra Datenbank & Web-Interface**

Datum der Ersteingabe: 23.09.2004.

Revisionsdatum: 16.01.2008.

Name der Institution: BGBM.

Verantwortlicher für das IT-Produkt: Dr. Regine Jahn [r.jahn@bgbm.org](mailto:r.jahn@bgbm.org); Anton Güntsch, [a.guentsch@bgbm.org](mailto:a.guentsch@bgbm.org).

Beteiligte Softwareentwickler: Botanischer Garten und Botanisches Museum, Berlin-Dahlem, Freie Universität Berlin, Biodiversitätsinformatik, Kontakt: Anton Güntsch, [a.guentsch@bgbm.org](mailto:a.guentsch@bgbm.org).

Die IT-Entwicklung ist für Dritte verfügbar (Datenbank verfügbar) oder nicht verfügbar. Lizenztyp: Lizenzpolitik ungeklärt.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Synonyme, Sammlungsbelege (Naturgeschichtliche Sammlungen), Beobachtungsdaten (Monitoring, Erfassungen), Bilddaten, geographische Daten, molekulare Daten, ökologische Daten, beschreibende Daten, Literaturverwaltung und Linksammlung.

Auf das IT-Produkt kann ausschließlich über das WWW zugegriffen werden.

Art des IT-Produktes: Datenbankanwendung und Web-Schnittstelle.

Kurze Beschreibung: Datenbank über Mikroalgen.

URL für das IT-Produkt: <http://www.algaterra.org>, <http://www.algaterra.net>.

Das IT-Produkt ist modularisiert, läuft als stabile Version. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar. Eine technische Dokumentation ist in Bearbeitung. Programmiersprache(n) des IT-Produkts: andere. Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar.

Die Quelltexte sind in englischer Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: MS SQL Server. Ein publiziertes Datenmodell existiert. Bezeichnung des Datenmodells: Berlin-Modell.

Datenbank Import-Format: Feldbegrenzer-limitierter Text. Datenbank Export-Format: XML.

"Authority files" wurden integriert. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist nicht eingeschränkt. Die Datenbankinhalte können online nur von bestimmten Usern editiert werden und nur in bestimmten Bereichen editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC. Realisierung (Technologie) der Web Schnittstelle: Java Applets. Programmiersprache für Datenbank-Web Schnittstelle: ColdFusion.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? keine Einschränkung. Welche Laufzeitumgebung ist für das IT-Produkt notwendig? keine Einschränkung. Verwaltung des Web Service über: UDDI. Interface für Wrapper: BioCASE.

### **2 BIG (Bundesinformationssystem Genetische Ressourcen): Teilprojekt Wildpflanzen**

Datum der Ersteingabe: 27.9.2004.

Revisionsdatum: 24.04.07.

Name der Institution: Bundesamt für Naturschutz im Verbund mit anderen Institutionen (RUB, IPK, ZADI).

Verantwortlicher für das IT-Produkt: Dr. Hans G. Fink (finkhg@bfn.de), Koordinator Siegfried Harrer (ZADI).

Beteiligte Softwareentwickler: Firma doctronic GmbH & Co.KG: (info@doctronic.de), Rudolf May (rudolf.may@bfn.de), Herbert Weber (weberh@bfn.de).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: Lizenzpolitik ungeklärt.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Synonyme, Beobachtungsdaten (Monitoring, Erfassungen), Bilddaten, geographische Daten, ökologische Daten, Schutzstatus, beschreibende Daten und Literaturverwaltung. Das IT-Produkt kann ausschließlich über das Internet genutzt werden.

Art des IT-Produktes: Web-Schnittstelle und Webservice, Wrapper.

Kurze Beschreibung: Ziel des Projektes ist die Entwicklung eines Informationssystems, das biologische, genetische, ökologische, ökonomische und geographische Informationen in einem System abbildet, um dadurch Synergieeffekte zu erzeugen und vernetzte Abfragen zu ermöglichen. Dabei sollen Techniken entwickelt werden, die heterogene, dezentrale Datenbanken durch einen zentralen Zugang in einem effizienten Online-Informationssystem verbinden, um sie dezentral abrufbar anzubieten. Es wird ein nutzerfreundliches Navigationssystem entwickelt, das den Einstieg über verschiedenartige Kriterien wie Pflanzennamen, züchterisch relevante Merkmale (Werteigenschaften), geographische Referenzen u.ä. (Faktendaten) erlaubt. URL für das IT-Produkt: <http://www.big-flora.de>.

Das IT-Produkt ist modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt läuft als stabile Version. Die in der Benutzeroberfläche verwendete Sprache ist mehrsprachig (deutsch/englisch). Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar. Eine technische Dokumentation ist nicht verfügbar. Programmiersprache(n) des IT-Produkts: Java und andere. Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar.

Die Quelltexte sind in deutscher Sprache kommentiert.

### **3 Biodat**

Datum der Ersteinreichung: 21.04.2004.

Revisionsdatum: 11.07.2007.

Name der Institution: Museum Alexander Koenig/Bonn, Museum für Naturkunde/Berlin.

Verantwortlicher für das IT-Produkt: Dirk Striebing (dirk.striebling@museum.hu-berlin.de).

Beteiligte Softwareentwickler: Dr. Karl-Heinz Lampe (k.lampe.zfmk@uni-bonn.de).

Die IT-Entwicklung ist für Dritte nicht verfügbar. Lizenztyp: Produkt für andere Institutionen nicht verfügbar.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei Dr. Karl-Heinz Lampe & Dirk Striebing, Museum Koenig/Bonn.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Synonyme, Sammlungsbelege (Naturgeschichtliche Sammlungen), Bilddaten, geographische Daten, Daten zur Präparation und Aufbewahrung und Linksammlung.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden.

Art des IT-Produktes: Datenbankanwendung, Client-Software, Anwenderprogramm, Web-Schnittstelle und Webservice, Wrapper (BioCASE-wrapper). URL für das IT-Produkt: <http://www.biodat.de>.

Das IT-Produkt ist modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache und Microsoft IIS.

Das IT-Produkt läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist mehrsprachig (deutsch und englisch). Ein Hilfesystem oder eine Benutzeranleitung ist nicht verfügbar. Eine technische Dokumentation ist nicht verfügbar. Programmiersprache(n) des IT-Produkts: VB und andere (Pascal).

Entwicklungsumgebung(en): Borland Dephi 7.

Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar. Die Quelltexte sind in deutscher oder englischer Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: MS SQL Server (2000 (version 8)). Ein publiziertes Datenmodell existiert nicht.

Datenbank Import-Format: Feldbegrenzer-limitierter Text. Datenbank Export-Format: XML und Feldbegrenzer-limitierter Text. Datenbank-Datenbank-Content-Transfer über: ODBC oder andere (ADO, MS-OLEDB, BDE).

"Authority files" wurden nicht integriert. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist eingeschränkt (sensible Fundorte/Datensätze, Aufenthaltsort in der Sammlung). Die Datenbankinhalte können online nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC oder andere (ADO, BDE). Realisierung (Technologie) der Web Schnittstelle: ASP. Programmiersprache für Datenbank-Web Schnittstelle: JavaScript und VBScript.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Win32. Interface für Wrapper: BioCASE.

#### **4 BIOTABase-Family**

Datum der Ersteintragung: 28.02.2007.

Name der Institution: Universität Hamburg Biozentrum Klein Flottbek & Botanischer Garten.

Verantwortlicher für das IT-Produkt: Dr. Gerhard Muehe (gerhard.muehe@botanik.uni-hamburg.de).

Beteiligte Softwareentwickler: Dipl. Ing. Andrzej Suwald (andrzej.suwald@botanik.uni-hamburg.de).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: andere Lizenz.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei Dr. G. Muehe, Prof. Dr. N. Jürgens.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Synonyme, Sammlungsbelege (Naturgeschichtliche Sammlungen), Beobachtungsdaten (Monitoring, Erfassungen), Bilddaten, Audio-Daten, meteorologische Daten, molekulare Daten, Populationsdaten, ökologische Daten, Schutzstatus und beschreibende Daten.

Das IT-Produkt kann lokal genutzt werden.

Art des IT-Produktes: Datenbankanwendung und Anwenderprogramm.

Kurze Beschreibung: Die Produkte der BIOTABase-Family dienen der systematischen Erfassung, Verwaltung und Auswertung von Monitoringdaten zu Biodiversität nach der Methodik von BIOTA AFRICA. Derzeitigen Schwerpunkt bilden die Vegetationsdaten des südlichen und nördlichen Afrikas. Abiotische sowie Daten zur Landnutzung werden im Kontext mitverwaltet. Die Aufnahme weiterer Datenbestände, sowohl nach Menge als auch nach Struktur ist in kontinuierlicher Entwicklung. URL für das IT-Produkt: <http://www.biota-africa.org>.

Das IT-Produkt ist modularisiert, läuft als Beta Testversion (Version 20060828). Die in der Benutzeroberfläche verwendete Sprache ist englisch und andere. Ein Hilfesystem oder eine Benutzeranleitung ist offline verfügbar. Eine technische Dokumentation ist offline verfügbar. Programmiersprache(n) des IT-Produkts: C/C++.

Entwicklungsumgebung(en): Visual C++6.0, MFC.

Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar. Die Quelltexte sind in deutscher Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: MS Access (Version 97). Ein publiziertes Datenmodell ist in Bearbeitung.

Datenbank Import-Format: Feldbegrenzer-limitierter Text. Datenbank Export-Format: Feldbegrenzer-limitierter Text und andere.

"Authority files" wurden nicht integriert. Eine Datenbankabfrage online ist nicht möglich.



Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Windows 98 aufwärts.  
Welche Laufzeitumgebung ist für das IT-Produkt notwendig? Windows 98 aufwärts. Bisher wurde das IT-Produkt extern genutzt durch BIOTA AFRICA (Modul soil science).

## **5 Bogart**

Datum der Ersteingabe: 30.9.2004.

Revisionsdatum: 30.6.2006.

Name der Institution: Botanischer Garten und Botanisches Museum Berlin (BGBM).

Verantwortlicher für das IT-Produkt: Anton Güntsch.

Beteiligte Softwareentwickler: Anton Güntsch (a.guentsch@bgbm.org), Frank Specht (f.specht@bgbm.org), Dominik Röpert (d.roepert@bgbm.org).

Die IT-Entwicklung ist für Dritte verfügbar (ohne Support). Lizenztyp: Lizenzpolitik ungeklärt.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Synonyme, Sammlungsbelege (Naturgeschichtliche Sammlungen), Bilddaten, geographische Daten, Schutzstatus, Daten zur Präparation und Aufbewahrung und Literaturverwaltung.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden.

Art des IT-Produktes: Datenbankanwendung.

Kurze Beschreibung: Akzessionsdatenbank System, mit dem die Lebendsammlung und das Gartenherbarium des BGBM vollständig verwaltet wird. URL für das IT-Produkt: <http://ww2.bgbm.org/bogartdb/BogartPublic.asp>.

Das IT-Produkt ist nicht modularisiert, läuft als stabile Version. Die in der Benutzeroberfläche verwendete Sprache ist deutsch. Ein Hilfesystem oder eine Benutzeranleitung ist offline verfügbar. Eine technische Dokumentation ist offline verfügbar. Programmiersprache(n) des IT-Produktes: VB. Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar.

Die Quelltexte sind in deutscher Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: MS SQL Server. Ein publiziertes Datenmodell existiert.

Datenbank Import-Format: andere. Datenbank Export-Format: andere.

Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist nicht eingeschränkt. Die Datenbankinhalte können online nicht editiert werden. Realisierung (Technologie) der Web Schnittstelle: ASP.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Windows NT oder XP. Inventursoftware für mobile Endgeräte.

## **6 CeDaMar-Webinterface**

Datum der Ersteingabe: 28.9.04.

Name der Institution: FIS, Abteilung Deutsches Zentrum für Marine Biodiversitätsforschung (DZMB).

Verantwortlicher für das IT-Produkt: Burkhard Küster, [bkuester@senckenberg.de](mailto:bkuester@senckenberg.de).

Art des IT-Produktes: Web-Schnittstelle. Das IT-Produkt läuft als stabile Version.

## **7 DALI - Database of Ascomycete Literature**

Datum der Ersteingabe: 04.05.2004.

Revisionsdatum: 19.07.2007.

Name der Institution: Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT-Zentrum.

Verantwortlicher für das IT-Produkt: Dr. Dagmar Triebel [triebeld@bsm.mwn.de](mailto:triebeld@bsm.mwn.de).

Beteiligte Softwareentwickler: Wiltrud Spiesberger ([spiesberger@bms.mwn.de](mailto:spiesberger@bms.mwn.de)), Jan Ingenhaag ([ingenhaag@bms.mwn.de](mailto:ingenhaag@bms.mwn.de)).

Die IT-Entwicklung ist für Dritte verfügbar.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: beschreibende Daten.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden oder lokal genutzt werden.

Art des IT-Produktes: Datenbankanwendung und Web-Schnittstelle.

Kurze Beschreibung: Access-Datenbank, ColdFusion-Webinterface. URL für das IT-Produkt: <http://dali.botanischestaatssammlung.de/>.

Das IT-Produkt ist nicht modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar. Eine technische Dokumentation ist nicht verfügbar. Programmiersprache(n) des IT-Produkts: VB. Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar. Die Quelltexte sind nicht kommentiert.

Verwendetes Datenbank-Verwaltungssystem: MS Access. Ein publiziertes Datenmodell existiert nicht. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist eingeschränkt. Die Datenbankinhalte können online nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC. Realisierung (Technologie) der Web Schnittstelle: andere (ColdFusion). Programmiersprache für Datenbank-Web Schnittstelle: ColdFusion.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Webinterface systemunabhängig, Datenbank unter Windows (NT, 2000, XP). Bisher wurde das IT-Produkt extern genutzt durch Web-Nutzer.

Geplante Erweiterungen: Umstellung auf DiversityReferences (siehe <http://www.diversityworkbench.net>).

## **8 Datenbank Herbarium Hamburgense**

Datum der Ersteintragung: 22.9.2004.

Revisionsdatum: 20.02.2007.

Name der Institution: Universität Hamburg Biozentrum Klein Flottbek & Botanischer Garten.

Verantwortlicher für das IT-Produkt: Dr. Gerhard Muche ([gerhard.muche@botanik.uni-hamburg.de](mailto:gerhard.muche@botanik.uni-hamburg.de)).

Beteiligte Softwareentwickler: Dipl. Chem. Thomas Hillmann ([thomas.hillmann@botanik.uni-hamburg.de](mailto:thomas.hillmann@botanik.uni-hamburg.de)).

Die IT-Entwicklung ist für Dritte verfügbar (auf Anfrage). Lizenztyp: andere Lizenz.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Sammlungsbelege (Naturgeschichtliche Sammlungen), Bilddaten, geographische Daten und Schutzstatus.

Das IT-Produkt kann ausschließlich über das Internet genutzt werden.

Art des IT-Produktes: Datenbankanwendung und Web-Schnittstelle. URL für das IT-Produkt: <http://www.HerbariumHamburgense.uni-hamburg.de>.

Das IT-Produkt ist nicht modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist nicht verfügbar und online verfügbar (intuitive Benutzerführung). Eine technische Dokumentation ist nicht verfügbar. Programmiersprache(n) des IT-Produkts: andere und C/C++ (PHP5). Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar. Die Quelltexte sind in deutscher Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: MySQL oder MS Access (vorläufig, bzw. als Zwischenlösung). Ein publiziertes Datenmodell existiert nicht.

Datenbank Export-Format: Feldbegrenzer-limitierter Text. Datenbank-Datenbank-Content-Transfer über: andere.

"Authority files" wurden nicht integriert. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist eingeschränkt (Fundstellen geschützter Arten ausgeblendet). Die Datenbankinhalte können online nur von bestimmten Usern editiert werden und nur in bestimmten Bereichen editiert werden. Programmiersprache für Datenbank-Web Schnittstelle: PHP.

Interface für Wrapper: BioCASE.

## **9 Digitale Belege**

Datum der Ersteingabe: 30.9.2004.

Name der Institution: Botanischer Garten und Botanisches Museum Berlin (BGBM).

Verantwortlicher für das IT-Produkt: Dominik Röpert.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Bilddaten und geographische Daten.

Art des IT-Produktes: Datenbankanwendung.

Das IT-Produkt läuft als Beta Testversion.

## **10 DiversityCollection**

Datum der Ersteingabe: 15.04.2004.

Revisionsdatum: 19.07.2007.

Name der Institution: Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT Zentrum.

Verantwortlicher für das IT-Produkt: Dr. Dagmar Triebel (triebel@bsm.mwn.de), Dr. Markus Weiss.

Beteiligte Softwareentwickler: Dr. Markus Weiss (weiss@bsm.mwn.de).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: OSI-zertifizierte open source Software. Bezeichnung der Lizenz: GPL.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: Synonyme und Bilddaten.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden.

Art des IT-Produktes: Datenbankanwendung und Client-Software.

Kurze Beschreibung: Verwaltung naturwissenschaftlicher Sammlungsdaten. URL für das IT-Produkt: <http://www.diversityworkbench.net>.

Das IT-Produkt ist modularisiert (DiversityWorkbench).

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar und in Bearbeitung (neue Version). Eine technische Dokumentation ist online verfügbar oder in Bearbeitung (Neue Version). Programmiersprache(n) des IT-Produkts: .Net (C#).

Entwicklungsumgebung(en): Microsoft Visual Studio .Net.

Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar (auf Nachfrage). Die Quelltexte sind in englischer Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: MS SQL Server. Ein publiziertes Datenmodell existiert.

Datenbank Import-Format: XML und andere. Datenbank Export-Format: XML und andere. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien oder ODBC.

"Authority files" wurden integriert. Welche "authority files" können verwendet werden? Andere Module der Diversity Workbench. Web Services via Index Fungorum-Web Service. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist eingeschränkt oder nicht eingeschränkt (über ColdFusion-Schnittstellen rein lesend für alle Internetnutzer). Die Datenbankinhalte können online nur von bestimmten Usern editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): .Net. Realisierung (Technologie) der Web Schnittstelle: andere (ColdFusion). Programmiersprache für Datenbank-Web Schnittstelle: ColdFusion und .Net.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Windows. Welche Laufzeitumgebung ist für das IT-Produkt notwendig? .Net Framework 2.0. Interface für Wrapper: Bio-CASE. Bisher wurde das IT-Produkt extern genutzt durch L. Zedda, Universität Bonn, H. Boyle, Naturkundemuseum Görlitz, Y. Novozilov, Komorov Institut, St. Petersburg, API-Projekt: Herbarien in M, HOH, MSB, REG, STU, TUB, H. Sipman, BGBM Berlin.

## 11 DiversityExsiccatae

Datum der Ersteingabe: 06.05.2004.

Revisionsdatum: 19.07.2007.

Name der Institution: Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT-Zentrum.

Verantwortlicher für das IT-Produkt: Dr. Dagmar Triebel (triebel@bsm.mwn.de), Dr. Markus Weiss.

Beteiligte Softwareentwickler: Gregor Hagedorn (g.hagedorn@bba.de), Wiltrud Spiesberger (Webinterface), Dr. Markus Weiss (weiss@bsm.mwn.de) (DiversityExsiccatae), Asif R. Lakehsar (lakehsar@bsm.mwn.de) (Webservice, Webinterface).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: andere Lizenz.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: beschreibende Daten (Bibliographie von botanischen und mykologischen Exsikkatenwerken) und Literaturverwaltung.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden oder lokal genutzt werden.

Art des IT-Produktes: Datenbankanwendung, Client-Software, Web-Schnittstelle und Webservice, Wrapper (Axis-2 Webservice ab Herbst 2007).

Kurze Beschreibung:	Datenbankapplikation	DiversityExsiccatae	siehe
<a href="http://www.diversityworkbench.net">http://www.diversityworkbench.net</a> ,	Projekt: IndExs -	Index of Exsiccatae,	siehe
<a href="http://indexs.botanischestaatssammlung.de/">http://indexs.botanischestaatssammlung.de/</a> .	URL	für das	IT-Produkt:
<a href="http://indexs.botanischestaatssammlung.de/">http://indexs.botanischestaatssammlung.de/</a> .			

Das IT-Produkt ist nicht modularisiert (wird allerdings parallel als Modul in DiversityCollection verwendet).

Für IT-Produkt mögliche HTTP Server: Apache und Microsoft IIS.

Das IT-Produkt läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist nicht verfügbar. Eine technische Dokumentation ist z. T. online verfügbar. Programmiersprache(n) des IT-Produkts: .Net. Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar. Die Quelltexte sind nicht kommentiert.

Verwendetes Datenbankmanagementsystem: MS SQL Server. Ein publiziertes Datenmodell existiert. Bezeichnung des Datenmodells: siehe <http://www.diversityworkbench.net>.

Datenbank Import-Format: XML. Datenbank Export-Format: XML. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien.

Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist eingeschränkt. Die Datenbankinhalte können online nur von bestimmten Usern editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): JDBC, .Net oder andere. Realisierung (Technologie) der Web Schnittstelle: JSP und andere (ColdFusion). Programmiersprache für Datenbank-Web Schnittstelle: ColdFusion, Java und .Net.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Windows für .Net Client, sonst beliebig. Verbindungsprotokoll für Webservices: SOAP. Bisher wurde das IT-Produkt extern durch als in DiversityCollection integriertes Modul genutzt von: GBIF-D-Mykologie-Teilprojekten zur Sammlungserfassung, ansonsten Webnutzer.

Geplante Erweiterungen: Ausbau des Web Services.

## 12 DiversityNavigator

Datum der Ersteintragung: 15.11.04.

Revisionsdatum: 27.2.2007.

Name der Institution: Universität Bayreuth.

Verantwortlicher für das IT-Produkt: Luciana Zedda, Gerhard Rambold.

Beteiligte Softwareentwickler: Dieter Neubacher.

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: OSI-zertifizierte open source Software. Bezeichnung der Lizenz: GPL.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: Sammlungsbelege (Naturgeschichtliche Sammlungen), Beobachtungsdaten (Monitoring, Erfassungen) und ökologische Daten.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden.

Art des IT-Produktes: Client-Software. URL für das IT-Produkt: <http://www.diversitynavigator.net>.

Das IT-Produkt läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar. Eine technische Dokumentation ist online verfügbar. Programmiersprache(n) des IT-Produkts: Java. Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar (auf Anfrage).

Welche Laufzeitumgebung ist für das IT-Produkt notwendig? Java™ 2 Runtime Environment.

## 13 DiversityTaxonNames

Datum der Ersteintragung: 19.07.2007.

Name der Institution: Staatliche Naturwissenschaftliche Sammlungen Bayerns.

Verantwortlicher für das IT-Produkt: Dr. Dagmar Triebel ([triebel@bsm.mwn.de](mailto:triebel@bsm.mwn.de)), Dr. Markus Weiss.

Beteiligte Softwareentwickler: Dr. Markus Weiss ([weiss@bsm.mwn.de](mailto:weiss@bsm.mwn.de)).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: OSI-zertifizierte open source Software oder andere Lizenz. Bezeichnung der Lizenz: GPL.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden.

Art des IT-Produktes: Datenbankanwendung, Client-Software und Web-Schnittstelle.

Kurze Beschreibung: A database application for managing taxon hierarchy, synonymy and nomenclatural information. URL für das IT-Produkt: <http://www.diversityworkbench.net>.

Das IT-Produkt ist nicht modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache und Microsoft IIS.

Das IT-Produkt läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar. Eine technische Dokumentation ist online verfügbar. Programmiersprache(n) des IT-Produkts: VB und .Net.

Entwicklungsumgebung(en): .Net Framework 2.0.

Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar (auf Anfrage). Verwendetes Datenbank-Verwaltungssystem: MS SQL Server. Ein publiziertes Datenmodell existiert. Bezeichnung des Datenmodells: siehe unter <http://www.diversityworkbench.net>.

Datenbank Import-Format: XML und andere. Datenbank Export-Format: XML und andere. Datenbank-Datenbank-Content-Transfer über: JDBC.

Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist eingeschränkt. Die Datenbankinhalte können online nur von bestimmten Usern editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC oder .Net. Realisierung (Technologie) der Web Schnittstelle: JSP, .Net und andere. Programmiersprache für Datenbank-Web Schnittstelle: ColdFusion und Java.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? .Net Client: MS Windows, ansonsten beliebig. Verbindungsprotokoll für Webservices: SOAP. Bisher wurde das IT-Produkt extern genutzt durch Nutzer zur Dateneingabe wissenschaftlicher Namen von Myxomycetes, Ery-siphales, Flechten, Blütenpflanzen (Melastomataceae) und Vertebraten.

Geplante Erweiterungen: Schnittstellen zu DiversityReferences und DiversityCollection, Axis 2 Web-Service.

#### **14 DiversityWorkbench, bestehend aus Komponenten: DiversityCollection, DiversityDescriptions, DiversityReferences, DiversityResources, DiversityIndexing, DiversityTaxonNames, DiversityUsers, DiversityGazetteer etc.**

Datum der Ersteintragung: 30.4.2004.

Revisionsdatum: 20.6.2006.

Name der Institution: Biologische Bundesanstalt für Land- und Forstwirtschaft.

Verantwortlicher für das IT-Produkt: s.u. einzelne Modulen.

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: OSI-zertifizierte open source Software. Bezeichnung der Lizenz: GPL.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei Gregor Hagedorn & al., je nach Komponente.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Synonyme, Sammlungsbelege (Naturgeschichtliche Sammlungen), Beobachtungsdaten (Monitoring, Erfassungen), ökologische Daten, beschreibende Daten und Daten zur Präparation und Aufbewahrung.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden, lokal genutzt werden oder über das Internet (sowohl lokal als auch über http).

Art des IT-Produktes: Datenbankanwendung. URL für das IT-Produkt: <http://www.diversityworkbench.net>.

Das IT-Produkt ist modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt in Entwicklung (je nach Komponente) oder läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist in Bearbeitung. Eine technische Dokumentation ist online verfügbar. Programmiersprache(n) des IT-Produkts: VB, .Net und andere (ColdFusion (Entwicklung läuft aus)). Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar. Die Quelltexte sind in englischer Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: MS SQL Server oder MS Access (Access XP (=10)). Ein publiziertes Datenmodell existiert (siehe Modelle unter <http://www.diversityworkbench.net>). Bezeichnung des Datenmodells: DiversityWorkbench, bestehend aus Komponenten: DiversityCollection, DiversityDescriptions, DiversityReferences, DiversityResources, DiversityIndexing, DiversityTaxonNames, DiversityUsers, DiversityGazetteer etc.

Datenbank Import-Format: XML, Feldbegrenzer-limitierter Text und andere. Datenbank Export-Format: XML, Feldbegrenzer-limitierter Text und andere. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien, ODBC oder JDBC.

"Authority files" wurden integriert (DiversityGazetteer, based on Getty Thesaurus und Bundesamt für Geodäsie und Kartographie-Daten). Eine Datenbankabfrage online ist mit statischer Dokumentdarstellung möglich, mit dynamischer Dokumentengenerierung möglich oder in Bearbeitung. Die Datenbankinhalte können online noch nicht editiert werden. Realisierung (Technologie) der Web Schnittstelle: ASP, .Net und andere (ColdFusion (läuft aus)). Programmiersprache für Datenbank-Web Schnittstelle: Python/Zope und andere.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Windows 32 bit, alle Versionen, Web Produkte betriebssystemunabhängig. Welche Laufzeitumgebung ist für das IT-Produkt notwendig? Z. T. Net Framework.

## 15 EmailDirectory

Datum der Ersteingabe: 06.05.2004.

Revisionsdatum: 19. 07.2007.

Name der Institution: Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT-Zentrum.

Verantwortlicher für das IT-Produkt: Dr. Dagmar Triebel (triebel@bsm.mwn.de).

Beteiligte Softwareentwickler: Dieter Neubacher (neubacher@bsm.mwn.de), Wiltrud Spiesberger.

Die IT-Entwicklung ist für Dritte nicht verfügbar.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: Literaturverwaltung (Email-Adressen).

Das IT-Produkt kann über Client-Server Anbindung genutzt werden (Abfrage über Internet, Eingabe über DiversityNavigator).

Art des IT-Produktes: Datenbankanwendung und Web-Schnittstelle.

Kurze Beschreibung: Sammlung von Email-Adressen von Mykologen und Phykologen weltweit. URL für das IT-Produkt: [http://www.mycology.net/Contacts/EmailDirectory/MNet\\_index.cfm](http://www.mycology.net/Contacts/EmailDirectory/MNet_index.cfm), [http://www.phycology.net/Contacts/EmailDirectory/PNet\\_index.cfm](http://www.phycology.net/Contacts/EmailDirectory/PNet_index.cfm).

Das IT-Produkt ist nicht modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache und Microsoft IIS.

Das IT-Produkt läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist nicht verfügbar. Eine technische Dokumentation ist nicht verfügbar. Verwendetes Datenbank-Verwaltungssystem: PostgreSQL. Ein publiziertes Datenmodell existiert nicht.

Datenbank Import-Format: XML. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien.

"Authority files" wurden nicht integriert. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist nicht eingeschränkt. Die Datenbankinhalte können online global editiert werden (nur Eingabe von neuen Datensätzen, keine Möglichkeit zum Ändern oder Löschen von Datensätzen). Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC. Realisierung (Technologie) der Web Schnittstelle: andere (ColdFusion). Programmiersprache für Datenbank-Web Schnittstelle: ColdFusion.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Betriebssystemunabhängig. Bisher wurde das IT-Produkt extern genutzt durch Internet-Nutzer aus dem Bereich Mykologie und Phykologie.

## 16 FloraWeb

Datum der Ersteingabe: 27.9.2004.

Revisionsdatum: 27.04.2007.

Name der Institution: Bundesamt für Naturschutz.

Verantwortlicher für das IT-Produkt: Rudolf May.

Beteiligte Softwareentwickler: doctronic GmbH (kueper@doctronic.de), Fraunhofer Autonomous Intelligent Systems Institute – AIS (mark.ostrovskii@ais.fraunhofer.de).

Die IT-Entwicklung ist für Dritte verfügbar (unter <http://www.floraweb.de>). Lizenztyp: Lizenzpolitik ungeklärt.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Synonyme, Beobachtungsdaten (Monitoring, Erfassungen), Bilddaten, geographische Daten, Populationsdaten, ökologische Daten, Schutzstatus, beschreibende Daten, Literaturverwaltung, Linksammlung und andere (Vegetationsaufnahmen).

Das IT-Produkt kann ausschließlich über das Internet genutzt werden.

Art des IT-Produktes: Datenbankanwendung, Anwenderprogramm, Web-Schnittstelle und Webservice, Wrapper.

Kurze Beschreibung: <http://www.floraweb.de> enthält ca. 50 Informationen einschl. Verbreitungskarten und Fotos zu allen heimischen Blütenpflanzenarten sowie Vegetationsaufnahmen, Pflanzengesellschaften und Vegetationskarten. Die Daten sind u.a. über GIS-Anwendungen auswertbar. URL für das IT-Produkt: <http://www.floraweb.de>.

Das IT-Produkt ist modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache und andere (Tomcat).

Das IT-Produkt läuft als stabile Version. Die in der Benutzeroberfläche verwendete Sprache ist deutsch. Ein Hilfesystem oder eine Benutzeranleitung ist in Bearbeitung. Eine technische Dokumentation ist offline verfügbar. Programmiersprache(n) des IT-Produkts: Java, VB und andere (Java Script, php).

Entwicklungsumgebung(en): PHP, Java.

Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar (z.T.). Die Quelltexte sind z. T. in deutscher Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: Oracle oder MySQL. Ein publiziertes Datenmodell existiert nicht.

Datenbank Import-Format: andere. Datenbank Export-Format: XML und andere. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien oder andere.

"Authority files" wurden integriert. Welche "authority files" können verwendet werden? Verschiedene geographische Referenzen (Kreise, Naturräume etc.). Taxonomische Referenz: Standardliste der Farn- und Blütenpflanzen. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist eingeschränkt (Teilbereiche nur für registrierte Anwender verfügbar). Die Datenbankinhalte können online noch nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): JDBC. Realisierung (Technologie) der Web Schnittstelle: CGI, Java servlets, Java applets und andere. Programmiersprache für Datenbank-Web Schnittstelle: PHP (Oracle-XSQL-Servlet), Java und andere.

Welche Laufzeitumgebung ist für das IT-Produkt notwendig? Webbrowser mit Javascript und Java Virtual Machine. Verwaltung des Web Service über: UDDI. Verbindungsprotokoll für Webservices: SOAP. Interface für Wrapper: BioCASE. Bisher wurde das IT-Produkt extern genutzt durch Naturschutzbehörden, Botanische Institute, ehrenamtliche Kartierer, fachlich interessierte Laien.

Geplante Erweiterungen: Verschiedenen Schnittstellen, z.B. zu GBIF international.

## 17 Foto Gallery Internet

Datum der Ersteinreichung: 22.9.2004.

Name der Institution: Universität Hamburg, BIOTA.

Verantwortlicher für das IT-Produkt: Dr. Gerhard Muehe.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Sammlungsbelege (Naturgeschichtliche Sammlungen), Bilddaten und geographische Daten.



Art des IT-Produktes: Web-Schnittstelle und Datenbankanwendung.

Das IT-Produkt ist in Planung.

## **18 Genbankinformationssystem (GBIS)**

Datum der Ersteintragung: 12.10.2004.

Revisionsdatum: 10.07.2007.

Name der Institution: Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) Gatersleben.

Verantwortlicher für das IT-Produkt: Dr. Helmut Knüpffer (knupffer@ipk-gatersleben.de).

Beteiligte Softwareentwickler: Markus Oppermann (opperman@ipk-gatersleben.de), Werner Schölch (schoelch@ipk-gatersleben.de).

Die IT-Entwicklung ist für Dritte nicht verfügbar (auf IPK-Genbank zugeschnittenes Produkt). Lizenztyp: Produkt für andere Institutionen nicht verfügbar.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei IPK Gatersleben.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Sammlungsbelege (Naturgeschichtliche Sammlungen), geographische Daten, beschreibende Daten und Literaturverwaltung.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden. Art des IT-Produktes: Datenbankanwendung, Client-Software, Anwenderprogramm, Web-Schnittstelle und Webservice, Wrapper.

Kurze Beschreibung: Akzessionsdaten (unit-level) zu ca. 150.000 Akzessionen pflanzengenetischer Ressourcen. GBIS/M - internes Sammlungsmanagementsystem; GBIS/I - Internetportal zur Recherche nach Akzessionen und Online-Saatgutbestellung. URL für das IT-Produkt: <http://www.ipk-gatersleben.de/Internet/Infrastruktur/Genbankinfosystem> (GBIS/I).

Das IT-Produkt ist modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt läuft als stabile Version. Die in der Benutzeroberfläche verwendete Sprache ist mehrsprachig (deutsch, englisch (nur Online-System GBIS/I)). Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar und in Bearbeitung. Eine technische Dokumentation ist offline verfügbar oder in Bearbeitung. Programmiersprache(n) des IT-Produkts: Java.

Entwicklungsumgebung(en): JDeveloper.

Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar. Die Quelltexte sind in deutscher oder englischer Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: Oracle (10g). Ein publiziertes Datenmodell existiert nicht (auf Anfrage verfügbar).

Datenbank Import-Format: Feldbegrenzer-limitierter Text. Datenbank Export-Format: XML, Feldbegrenzer-limitierter Text und andere. Datenbank-Datenbank-Content-Transfer über: JDBC.

"Authority files" wurden nicht integriert. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist eingeschränkt. Die Datenbankinhalte können online nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): JDBC oder JSP. Realisierung (Technologie) der Web Schnittstelle: JSP. Programmiersprache für Datenbank-Web Schnittstelle: Java.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? GBIS/I - beliebig (Webbrowser), GBIS/M - Windows. Welche Laufzeitumgebung ist für das IT-Produkt notwendig? JVM 1.5. Verwaltung des Web Service über: UDDI. Verbindungsprotokoll für Webservices: SOAP. Interface für Wrapper: BioCASE. Bisher wurde das IT-Produkt extern genutzt durch keine; für externe Nutzer nur Online-Zugriff. Web Services: div. Erweiterungsmodule für GBIS/M, Weiterentwicklung GBIS/I.

## 19 GLOBIS-D Webpräsentation

Datum der Ersteintragung: 11.10.2004.

Revisionsdatum: 25.04.2007.

Name der Institution: Staatliches Museum für Naturkunde Stuttgart. Verantwortlicher für das IT-Produkt: Norbert Hirneisen (nobbi@s2you.de).

Beteiligte Softwareentwickler: Aktuell: Norbert Hirneisen, Basisversion: Alexander Kroupa, Frank Streitig.

Die IT-Entwicklung ist für Dritte verfügbar (als Application Service). Lizenztyp: andere Lizenz (Bereitstellung als AppService gegen Wartungskostenerstattung).

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei Norbert Hirneisen.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Synonyme, Beobachtungsdaten (Monitoring, Erfassungen), Bilddaten, geographische Daten, molekulare Daten, Populationsdaten, ökologische Daten, beschreibende Daten, Daten zur Präparation und Aufbewahrung und Literaturverwaltung.

Das IT-Produkt kann lokal genutzt werden (lokale Filemaker-DB für Teile der Daten) oder über das Internet.

Art des IT-Produktes: Datenbankanwendung, Anwenderprogramm (Offline XML-basiertes Editoren-tool) und Web-Schnittstelle.

Kurze Beschreibung: Darstellung der Weltliste mehrerer Tagfalterfamilie mit Angaben zu wissenschaftlichen Namen, Typenverbleib, Vorkommen und Artsteckbriefen. URL für das IT-Produkt: <http://www.lepidat.org>.

Das IT-Produkt ist modularisiert (auf verteilten Servern mit Spezialisierung für Kartenausgabe, Datenbank, Bildverwaltung).

Für IT-Produkt mögliche HTTP Server: Apache (andere nicht getestet).

Das IT-Produkt läuft als Beta Testversion (Version 0.29). Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar. Eine technische Dokumentation ist offline verfügbar. Programmiersprache(n) des IT-Produkts: Java und Python (für Grafikmodule).

Entwicklungsumgebung(en): Eclipse. Webseitig: Struts (Java Framework).

Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar (auf Nachfrage). Die Quelltexte sind in deutscher oder englischer Sprache kommentiert (source-comments (spärlich) externe Dokumentation). Quelltext (CVS Account): auf Nachfrage.

Verwendetes Datenbank-Verwaltungssystem: Oracle oder Filemaker (Für die lokale Datenaufbereitung). Ein publiziertes Datenmodell existiert nicht (intern ER-Diagramm verfügbar).

Datenbank Import-Format: XML (Offline Autorentool) und Feldbegrenzer-limitierter Text. Datenbank Export-Format: XML (Export für Autoren) und Feldbegrenzer-limitierter Text. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien.

Welche "authority files" können verwendet werden? Das Projekt erstellt einen eigenen authority file für Tagfalter. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist eingeschränkt (Abfrage ist nur über die Applikation möglich). Die Datenbankinhalte können online nur von bestimmten Usern editiert werden (Registrierung ist nötig). Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): JSP oder andere. Realisierung (Technologie) der Web Schnittstelle: Java Servlets und JSP (über Struts-Framework). Programmiersprache für Datenbank-Web Schnittstelle: Java.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Ein Offline-Autorentool zur verteilten Bearbeitung der Artsteckbriefe ist in Arbeit. Der Datenaustausch erfolgt über eine XML-Webschnittstelle. Welche Laufzeitumgebung ist für das IT-Produkt notwendig? SUN mit solaris 8/9, JAVA 1.4.2+, Apache 1.3+, Tomcat 4.18 Andere Umgebungen wären wohl möglich, sind aber nicht getestet. Bildserver: python 2.3 mit PIL. Offline-Tool: Win32. Bisher wurde das IT-Produkt noch nicht freigegeben, und deswegen noch nicht extern genutzt.

Geplante Erweiterungen: Kartendarstellung, PDF-Ausgabe, automatisierter XML-Import.

## **20 Herbardatenbank des IPK Gatersleben**

Datum der Ersteingabe: 12.10.2004.

Revisionsdatum: 11.07.2007.

Name der Institution: Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) Gatersleben.

Verantwortlicher für das IT-Produkt: Dr. Helmut Knüpffer (knupffer@ipk-gatersleben.de).

Die IT-Entwicklung ist für Dritte nicht verfügbar.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Sammlungsbelege (Naturgeschichtliche Sammlungen), Bilddaten und geographische Daten.

Das IT-Produkt kann lokal genutzt werden (WWW-Anbindung über Mansfeld-Datenbank).

Art des IT-Produktes: Datenbankanwendung, Anwenderprogramm und Web-Schnittstelle.

Kurze Beschreibung: Bilddaten (Scan) und Etikettendaten von einigen tausend Belegen des IPK-Kulturpflanzenherbars. URL für das IT-Produkt: über Mansfeld-Datenbank <http://mansfeld.ipk-gatersleben.de/mansfeld/Query.htm>.

Das IT-Produkt ist nicht modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt in Entwicklung (Weiterentwicklung z.Z. zurückgestellt). Programmiersprache(n) des IT-Produktes: andere (Visual FoxPro). Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar. Verwendetes Datenbank-Verwaltungssystem: andere (Visual FoxPro). Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich oder in Bearbeitung (über Mansfeld-Datenbank). Die Datenbankinhalte können online nicht editiert werden. Informationen/Daten auf den Beleg-Etiketten sollen mit aufgenommen werden.

## **21 Interne MS Access Entwicklung**

Datum der Ersteingabe: 20.9.2004.

Name der Institution: Universität Bonn.

Verantwortlicher für das IT-Produkt: Dr. Jens Mutke.

Die IT-Entwicklung ist für Dritte nicht verfügbar.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Sammlungsbelege (Naturgeschichtliche Sammlungen), geographische Daten, geologische Daten, Populationsdaten und ökologische Daten.

Art des IT-Produktes: Datenbankanwendung (Datenmodell nach Publikation W. Berendsohn).

Das IT-Produkt läuft als Beta Testversion. Verwendetes Datenbank-Verwaltungssystem: MS Access. Ein publiziertes Datenmodell existiert (Publikation W. Berendsohn (Quelle)).

## **22 Internet Datenbank Biota-W**

Datum der Ersteingabe: 20.9.2004.

Name der Institution: Universität Bonn.

Verantwortlicher für das IT-Produkt: Dr. Jens Mutke.

Die IT-Entwicklung ist für Dritte nicht verfügbar.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Sammlungsbelege (Naturgeschichtliche Sammlungen), geographische Daten, geologische Daten, Populationsdaten und ökologische Daten.

Art des IT-Produktes: Web-Schnittstelle und Datenbankanwendung.

Das IT-Produkt läuft als Beta Testversion.

## **23 JsLit (Version 8, vormals 6)**

Datum der Ersteingabe: 22.9.2004.

Revisionsdatum: 12.07.2007.

Name der Institution: Zoologische Staatssammlung München (ZSM).

Verantwortlicher für das IT-Produkt: Jörg Spelda (spelda@zsm.mwn.de).

Die IT-Entwicklung ist für Dritte nicht verfügbar. Lizenztyp: Produkt für andere Institutionen nicht verfügbar.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei Jörg Spelda.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: Literaturverwaltung.

Das IT-Produkt kann lokal genutzt werden.

Art des IT-Produktes: Datenbankanwendung.

Kurze Beschreibung: JsLit verwaltet Zitate in einem zu ReferenceManager kompatiblen Format. Die Datenbank fungiert zudem als "literature backbone" für SysMyr und MyrKar. URL für das IT-Produkt: <http://www.gbifev2.mwn.de>.

Das IT-Produkt ist nicht modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt läuft als stabile Version. Die in der Benutzeroberfläche verwendete Sprache ist deutsch. Ein Hilfesystem oder eine Benutzeranleitung ist nicht verfügbar. Eine technische Dokumentation ist nicht verfügbar. Programmiersprache(n) des IT-Produkts: VB.

Entwicklungsumgebung(en): MS Access 2000.

Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar. Die Quelltexte sind nicht kommentiert.

Verwendetes Datenbank-Verwaltungssystem: MS Access. Ein publiziertes Datenmodell existiert. Bezeichnung des Datenmodells: siehe: SPELDA, J.; ROSENBERG, J.; VOIGTLÄNDER, K. (2003a): The German Myriapod Literature Project (GerMyLit). – African Invertebrates, 44 (1): 325-330. Pietermaritzburg.

Datenbank Import-Format: Feldbegrenzer-limitierter Text. Datenbank Export-Format: Feldbegrenzer-limitierter Text. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien oder ODBC.

Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich oder mit statischer Dokumentdarstellung möglich (ist Teil der SysMyr-Datenbankabfrage). Die Datenbankabfrage online ist eingeschränkt. Die Datenbankinhalte können online nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): JSP. Programmiersprache für Datenbank-Web Schnittstelle: PHP.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? WIN32 (ME, XP, 98, 2000 getestet). Welche Laufzeitumgebung ist für das IT-Produkt notwendig? MS Access 2000. Bisher wurde das IT-Produkt extern genutzt durch Arbeitsgemeinschaft deutschsprachiger Myriapodologen, Dr. Hans S. Reip: reip@myriapoden-info.de, Jena: <http://www.myriapoden-info.de/MyriaLit/index.html>.

Geplante Erweiterungen: Eigenständige Literaturabfrage über <http://www.gbifev2.mwn.de>. Derzeit wird die Webpräsenz von der Arbeitsgemeinschaft deutschsprachiger Myriapodologen (Dr. Hans S. Reip) realisiert.

## 24 LinkCollection

Datum der Ersteintragung: 06.05.2004.

Revisionsdatum: 19.07.2007.

Name der Institution: Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT Zentrum.

Verantwortlicher für das IT-Produkt: Dr. Dagmar Triebel.

Beteiligte Softwareentwickler: Jan Ingenhaag (ingenhaag@bsm.mwn.de), Wiltrud Spiesberger, Dr. Markus Weiß (weiss@bsm.mwn.de).

Die IT-Entwicklung ist für Dritte nicht verfügbar.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: Daten zur Präparation und Aufbewahrung.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden oder lokal genutzt werden.

Art des IT-Produktes: Datenbankanwendung und Web-Schnittstelle.

Kurze Beschreibung: Strukturierte Linksammlung für mykologische und phykologische Biodiversitätsdaten. URL für das IT-Produkt: <http://www.mycology.net>, <http://www.phycology.net>.

Das IT-Produkt ist nicht modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache und Microsoft IIS.

Das IT-Produkt läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist nicht verfügbar. Eine technische Dokumentation ist nicht verfügbar. Programmiersprache(n) des IT-Produktes: andere (PostgreSQL-Datenbank, online editierbar über ColdFusion). Verwendetes Datenbank-Verwaltungssystem: PostgreSQL. Ein publiziertes Datenmodell existiert nicht.

Datenbank Import-Format: XML. Datenbank Export-Format: XML. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien.

Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist nicht eingeschränkt. Die Datenbankinhalte können online nur von bestimmten Usern editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC. Realisierung (Technologie) der Web Schnittstelle: andere (ColdFusion). Programmiersprache für Datenbank-Web Schnittstelle: ColdFusion.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Betriebssystemunabhängig, Client über Browser.

## 25 Mansfeld's World Database of Agricultural and Horticultural Crops (Mansfeld-Datenbank)

Datum der Ersteintragung: 12.10.2004.

Revisionsdatum: 12.07.2007.

Name der Institution: Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) Gatersleben.

Verantwortlicher für das IT-Produkt: Dr. Helmut Knüpffer (knupffer@ipk-gatersleben.de).

Beteiligte Softwareentwickler: Dr. Ram Narang (narang@ipk-gatersleben.de).

Die IT-Entwicklung ist für Dritte nicht verfügbar.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei IPK Gatersleben.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Synonyme, Sammlungsbelege (Naturgeschichtliche Sammlungen), Bilddaten, geographische Daten, beschreibende Daten und Literaturverwaltung.

Auf das IT-Produkt kann ausschließlich über das Internet zugegriffen werden.

Art des IT-Produktes: Datenbankanwendung und Web-Schnittstelle.

Kurze Beschreibung: Die Mansfeld-Datenbank basiert auf der englischen Buchausgabe "Mansfeld's Encyclopedia of Agricultural and Horticultural Crops" (2001) und enthält Angaben zu ca. 6.100 kultivierten Pflanzenarten (mit Ausschluss von Forst- und Zierpflanzen). Die Angaben je Art umfassen Taxonomie und Nomenklatur, Volksnamen in vielen Sprachen, Verbreitung (wild und in Kultur), wildwachsende Verwandte, Domestikation, Nutzung und Kulturzweck, sowie Literaturangaben. Die "Rehm-Datenbank" (Datenbank für tropische und subtropische Nutzpflanzen von Prof. S. Rehm, Göttingen) ist (in deutscher Sprache) ebenfalls über die Mansfeld-Datenbank recherchierbar. URL für das IT-Produkt: <http://mansfeld.ipk-gatersleben.de>.

Das IT-Produkt ist nicht modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist mehrsprachig (englisch (Mansfeld), deutsch (Rehm)). Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar. Eine technische Dokumentation ist in Bearbeitung. Programmiersprache(n) des IT-Produkts: andere (Oracle Application Express Release 2.2 (PL/SQL)). Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar.

Die Quelltexte sind in deutscher oder englischer Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: Oracle (Oracle Database 10g, Release 2).

Datenbank Import-Format: XML und Feldbegrenzer-limitierter Text. Datenbank Export-Format: XML und Feldbegrenzer-limitierter Text. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien.

Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist nicht eingeschränkt. Die Datenbankinhalte können online nur von bestimmten Usern editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): andere (Oracle Application Express, Release 2.2). Realisierung (Technologie) der Web Schnittstelle: andere (Oracle Application Express, Release 2.2). Verschiedenen Sprachmodule.

## **26 Materialarchiv Forschungsschiff Meteor**

Datum der Ersteingabe: 28.9.04.

Name der Institution: FIS, Abteilung Deutsches Zentrum für Marine Biodiversitätsforschung (DZMB).

Verantwortlicher für das IT-Produkt: Burkhard Küster ([bkuester@senckenberg.de](mailto:bkuester@senckenberg.de)).

Die IT-Entwicklung ist für Dritte verfügbar (auf Anfrage).

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Bilddaten, geographische Daten und meteorologische Daten.

Art des IT-Produktes: Datenbankanwendung.

Das IT-Produkt läuft als Beta Testversion. Verwendetes Datenbank-Verwaltungssystem: PostgreSQL.

## **27 MS Access Anwendung für Prokaryota Knoten**

Datum der Ersteingabe: 20.9.2004.

Name der Institution: DSMZ.

Verantwortlicher für das IT-Produkt: Dr. Manfred Kracht, Dr. Dorothea Gleim, Andrejs Naumovs.

Beteiligte Softwareentwickler: Dr. Manfred Kracht, Dr. Dorothea Gleim, Andrejs Naumovs.

Die IT-Entwicklung ist für Dritte nicht verfügbar.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Sammlungsbelege (Naturgeschichtliche Sammlungen), geographische Daten, molekulare Daten, beschreibende Daten, Daten zur Präparation und Aufbewahrung, Literaturverwaltung, andere (taxonomische Daten: Synonyme, Identifizierung, History der Kulturen, Wachstumsbedingungen, Chemie, Substrat-

verwertung, praktische Anwendungen, Patente, verwaltungstechnische (interne) Daten) und Link-sammlung.

Das IT-Produkt kann lokal genutzt werden.

Art des IT-Produktes: Datenbankanwendung.

Das IT-Produkt läuft als Beta Testversion oder ist erst in Planung. Verwendetes Datenbank-Verwaltungssystem: MS Access.

Interface für Wrapper: BioCASE.

## **28 MyrChar (Version 1)**

Datum der Ersteingabe: 22.9.2004.

Revisionsdatum: 12.07.2007.

Name der Institution: Zoologische Staatssammlung München.

Verantwortlicher für das IT-Produkt: Jörg Spelda (spelda@t-online.de, spelda@zsm.mwn.de).

Die IT-Entwicklung ist für Dritte nicht verfügbar. Lizenztyp: Produkt für andere Institutionen nicht verfügbar.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei Jörg Spelda.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: beschreibende Daten.

Das IT-Produkt kann lokal genutzt werden. Art des IT-Produktes: Datenbankanwendung.

Kurze Beschreibung: MyrChar verwaltet Merkmale, getrennt nach nominalen und metrischen Daten und soll eine Identifikation von Taxa (bzw. einen einschränkenden Suchfilter) ermöglichen.

Das IT-Produkt ist modularisiert, läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist deutsch. Ein Hilfesystem oder eine Benutzeranleitung ist nicht verfügbar. Eine technische Dokumentation ist nicht verfügbar. Programmiersprache(n) des IT-Produkts: VB.

Entwicklungsumgebung(en): MS Access 2000.

Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar.

Sind die Quelltexte kommentiert? Nein.

Verwendetes Datenbank-Verwaltungssystem: MS Access. Ein publiziertes Datenmodell existiert nicht.

Datenbank Import-Format: Feldbegrenzer-limitierter Text. Datenbank Export-Format: Feldbegrenzer-limitierter Text. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien oder ODBC.

"Authority files" wurden nicht integriert. Eine Datenbankabfrage online ist nicht möglich. Die Datenbankinhalte können online nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? WIN32 (ME, XP, 2000 getestet). Welche Laufzeitumgebung ist für das IT-Produkt notwendig? MS Access 2000.

Das Projekt ruht, wird derzeit nicht weitergeführt.

## **29 MyrKar (Version 6)**

Datum der Ersteingabe: 22.9.2004.

Revisionsdatum: 12.07.2007.

Name der Institution: Zoologische Staatssammlung München (ZSM).

Verantwortlicher für das IT-Produkt: Jörg Spelda (spelda@t-online.de, spelda@zsm.mwn.de).

Beteiligte Softwareentwickler: Jörg Spelda (spelda@zsm.mwn.de).

Die IT-Entwicklung ist für Dritte nicht verfügbar. Lizenztyp: Produkt für andere Institutionen nicht verfügbar.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei Jörg Spelda.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Sammlungsbelege (Naturgeschichtliche Sammlungen), Beobachtungsdaten (Monitoring, Erfassungen), geographische Daten und ökologische Daten.

Das IT-Produkt kann lokal genutzt werden. Art des IT-Produktes: Datenbankanwendung.

Kurze Beschreibung: Kartierungsdatenbank mit separaten Modulen zur Erfassung von Sammlungsbelegen.

Das IT-Produkt ist modularisiert, läuft als stabile Version. Die in der Benutzeroberfläche verwendete Sprache ist deutsch. Ein Hilfesystem oder eine Benutzeranleitung ist nicht verfügbar. Eine technische Dokumentation ist nicht verfügbar. Programmiersprache(n) des IT-Produkts: VB.

Entwicklungsumgebung(en): MS Access 2000.

Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar.

Die Quelltexte sind nicht kommentiert.

Verwendetes Datenbank-Verwaltungssystem: MS Access. Ein publiziertes Datenmodell existiert nicht.

Datenbank Import-Format: Feldbegrenzer-limitierter Text. Datenbank Export-Format: Feldbegrenzer-limitierter Text. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien oder ODBC.

Eine Datenbankabfrage online ist nicht möglich. Die Datenbankinhalte können online nicht editiert werden.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? WIN32 (98, ME, 2000, XP getestet). Welche Laufzeitumgebung ist für das IT-Produkt notwendig? MS Access 2000. Bisher wurde das IT-Produkt extern genutzt, inzwischen nur noch eigene, private Nutzung. Keine Nutzung an der Institution.

### **30 NaviKey 4 – a Java applet and application for accessing descriptive data coded in DELTA format**

Datum der Ersteintragung: 19.07.2007.

Name der Institution: Universität Bayreuth, Arbeitsgruppe G. Rambold.

Verantwortlicher für das IT-Produkt: Prof. Dr. G. Rambold.

Beteiligte Softwareentwickler: Dieter Neubacher (neubacher@bsm.mwn.de).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: andere Lizenz. Bezeichnung der Lizenz: GPL.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: beschreibende Daten.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden, lokal genutzt werden oder über das Internet.

Art des IT-Produktes: Datenbankanwendung, Client-Software und Web-Schnittstelle.

Kurze Beschreibung: a Java applet and application for accessing descriptive data coded in DELTA format; tool for creating interactive identification keys. URL für das IT-Produkt: <http://www.navikey.net/>.

Das IT-Produkt ist modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar. Eine technische Dokumentation ist online verfügbar. Programmiersprache(n) des IT-Produkts: Java. Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar. Die Quelltexte sind in englischer Sprache kommentiert. Quelltext (URL): <http://www.navikey.net/>.



Verwendetes Datenbank-Verwaltungssystem: PostgreSQL oder andere. Ein publiziertes Datenmodell existiert. Bezeichnung des Datenmodells: DiversityDescriptions, siehe <http://www.diversityworkbench.net>. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist nicht eingeschränkt. Die Datenbankinhalte können online nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): JDBC oder andere. Realisierung (Technologie) der Web Schnittstelle: Java applets. Programmiersprache für Datenbank-Web Schnittstelle: Java.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Java Runtime Environment. Bisher wurde das IT-Produkt extern genutzt durch siehe <http://www.navikey.net/webprojects.html>.

Siehe: <http://www.navikey.net/>.

### **31 NeoFlora-Converter**

Datum der Ersteingabe: 27.9.2004.

Revisionsdatum: 21.10.04. Name der Institution: Bundesamt für Naturschutz.

Verantwortlicher für das IT-Produkt: Rudolf May ([rudolf.may@bfn.de](mailto:rudolf.may@bfn.de)). Beteiligte Softwareentwickler: doctronic GmbH, Carsten Oberscheid, ([oberscheid@doctronic.de](mailto:oberscheid@doctronic.de)).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: Lizenzpolitik ungeklärt.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei doctronic GmbH.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Synonyme, ökologische Daten, beschreibende Daten (Daten zum Thema "invasive alien species") und Linksammlung.

Das IT-Produkt kann lokal genutzt werden.

Art des IT-Produktes: Web-Schnittstelle.

Kurze Beschreibung: Der NeoFlora-Converter erstellt und publiziert Beiträge für das Webangebot NeoFlora. Dazu wird ein Beitrag in WinWord mit der NeoFlora-Dokumentvorlage geschrieben, in ein XML-Zwischenformat exportiert und über den Converter in eine HTML-Seite umgewandelt. URL für das IT-Produkt: <http://www.neophyten.de>.

Das IT-Produkt ist nicht modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt läuft als stabile Version. Die in der Benutzeroberfläche verwendete Sprache ist deutsch. Ein Hilfesystem oder eine Benutzeranleitung ist offline verfügbar. Eine technische Dokumentation ist offline verfügbar. Programmiersprache(n) des IT-Produkts: andere. Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar (auf Anfrage).

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Windows 95 oder höher. Bisher wurde das IT-Produkt extern genutzt durch Institut für Ökologie der TU Berlin/AG Neobita.

### **32 Portal Virtuelles Herbarium (Botanischer Knoten)**

Datum der Ersteingabe: 30.6.2006.

Name der Institution: Botanischer Garten & Botanisches Museum Berlin.

Verantwortlicher für das IT-Produkt: A. Güntsch, W. Berendsohn.

Beteiligte Softwareentwickler: Jörg Holetschek ([j.holetschek@bgbm.org](mailto:j.holetschek@bgbm.org)).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: andere Lizenz.

Art des IT-Produktes: Web-Schnittstelle.

Kurze Beschreibung: (in Entwicklung): Zugriffsschnittstelle zu den Datenbeständen angeschlossener heterogener verteilter Sammlungsdatenbanken (Wrapper-Technologie, s. <http://www.biocase.org>). Verwendung einer konzeptorientierten Taxondatenbank (Standardliste, BfN) zur Ausweitung von Suchanfragen auf verwandte Konzepte (derzeit: Synonyme). URL für das IT-Produkt: <http://www.gbif.de/botanik/datenabfrage>.

Das IT-Produkt ist modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache (derzeit Tomcat Standalone).

Das IT-Produkt in Entwicklung. Die in der Benutzeroberfläche verwendete Sprache ist mehrsprachig (Zugriff auf den Prototyp besteht derzeit über <http://ww2.biocase.org/abcdSimple/index.html> (englischsprachig) und <http://www.gbif.de/botanik/datenabfrage/default.html> (deutschsprachig)). Ein Hilfesystem oder eine Benutzeranleitung ist nicht verfügbar. Eine technische Dokumentation ist in Bearbeitung. Programmiersprache(n) des IT-Produkts: Java.

Entwicklungsumgebung(en): Windows2000 / Tomcat und Linux / Tomcat.

Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar (derzeit jedoch nur Prototyp). Die Quelltexte sind in englischer Sprache kommentiert. Quelltext (CVS Account): <http://ww3.bgbm.org/viewcvs/viewcvs.cgi/test/UnitJSP/>.

Verwendetes Datenbank-Verwaltungssystem: PostgreSQL (Metadaten) oder MS SQL Server.

### 33 PyWrapper

Datum der Ersteintragung: 15.04.2004.

Revisionsdatum: 30.6.2006.

Name der Institution: Botanischer Garten und Botanisches Museum Berlin-Dahlem, Abt. Biodiversitätssinformatik.

Verantwortlicher für das IT-Produkt: W.Berendsohn ([w.berendsohn@bgbm.org](mailto:w.berendsohn@bgbm.org)), A.Güntsche ([a.guentsch@bgbm.org](mailto:a.guentsch@bgbm.org)).

Beteiligte Softwareentwickler: Markus Döring ([m.doering@bgbm.org](mailto:m.doering@bgbm.org)).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: OSI-zertifizierte open source Software. Bezeichnung der Lizenz: Mozilla Public Licence.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei BGBM.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: Synonyme und Sammlungsbelege (Naturgeschichtliche Sammlungen).

Das IT-Produkt kann lokal genutzt werden. Art des IT-Produktes: Webservice, Wrapper.

Kurze Beschreibung: Generische Wrappersoftware zur Kommunikation mit beliebigen Datenbanken (betriebssystemunabhängig, datenstrukturunabhängig) auf Basis des BioCASE Protokolls. URL für das IT-Produkt: <http://www.biocase.org/dev/wrapper/>.

Das IT-Produkt ist modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache und Microsoft IIS.

Das IT-Produkt läuft als Beta Testversion. Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar und offline verfügbar. Eine technische Dokumentation ist online verfügbar oder offline verfügbar. Programmiersprache(n) des IT-Produkts: Python. Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar. Die Quelltexte sind in englischer Sprache kommentiert. Quelltext (URL): <http://ww2.biocase.org/websvn/>.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Getestet mit Linux, Win32, Unix, MacOS. Verwaltung des Web Service über: UDDI. Interface für Wrapper: BioCASE (Der Wrapper implementiert das BioCASE Protokoll). Bisher wurde das IT-Produkt extern genutzt durch zahlreiche Nutzer in Deutschland und Europa.

## 34 science4you online Monitoring

Datum der Ersteingabe: 11.10.2004.

Revisionsdatum: 25.04.2007.

Name der Institution: science + communication Norbert Hirneisen.

Verantwortlicher für das IT-Produkt: Norbert Hirneisen (nobbi@s2you.de).

Beteiligte Softwareentwickler: Norbert Hirneisen, Andreas Schmidt, Jörn Merkel, Konrad Hauke, Tobias Braun. An der Basisversion: Frank Streitzig, André Wolf, Alexander Kroupa.

Die IT-Entwicklung ist für Dritte verfügbar (als Application Service). Lizenztyp: andere Lizenz (Basis-Anwendung ist kostenfrei.).

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei Norbert Hirneisen.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Synonyme, Sammlungsbelege (Naturgeschichtliche Sammlungen), Beobachtungsdaten (Monitoring, Erfassungen), Bilddaten, Audio-Daten, geographische Daten, meteorologische Daten, ökologische Daten, Schutzstatus, beschreibende Daten, Daten zur Präparation und Aufbewahrung und Linksammlung.

Das IT-Produkt kann ausschließlich über das Internet genutzt werden.

Art des IT-Produktes: Datenbankanwendung und Web-Schnittstelle.

Kurze Beschreibung: Erfassung von Beobachtungen von Wanderfaltern. Sprache deutsch. Online-Formulare, sofortige Kartenausgabe. URL für das IT-Produkt: <http://www.science4you.org/platform/monitoring/home/index.do>.

Das IT-Produkt ist modularisiert (auf verteilten Servern mit Spezialisierung für Kartenausgabe, Datenbank, Bildverwaltung).

Für IT-Produkt mögliche HTTP Server: Apache (andere nicht getestet).

Das IT-Produkt läuft als stabile Version (Version 2.21). Die in der Benutzeroberfläche verwendete Sprache ist deutsch (englisch geplant). Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar. Eine technische Dokumentation ist offline verfügbar. Programmiersprache(n) des IT-Produkts: Java und Python (für Grafikmodule).

Entwicklungsumgebung(en): Eclipse. Webseitig: Struts (Javaframework).

Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar (auf Nachfrage). Die Quelltexte sind in deutscher oder englischer Sprache kommentiert (source-comments (spärlich), externe Dokumentation). Quelltext (CVS Account): auf Nachfrage.

Verwendetes Datenbank-Verwaltungssystem: Oracle. Ein publiziertes Datenmodell existiert nicht (intern ER-Diagramm verfügbar).

Datenbank Import-Format: XML und Feldbegrenzer-limitierter Text. Datenbank Export-Format: XML und Feldbegrenzer-limitierter Text (csv-Format zum direkten Excel-Import). Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien.

Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist eingeschränkt (Abfrage ist nur über die Applikation möglich). Die Datenbankinhalte können online nur von bestimmten Usern editiert werden (Registrierung ist nötig) und nur in bestimmten Bereichen editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): JDBC oder JSP. Realisierung (Technologie) der Web Schnittstelle: Java servlets und JSP (über struts-framework). Programmiersprache für Datenbank-Web Schnittstelle: Java.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? +/- aktueller Browser mit aktivierten Cookies und Javascript. Welche Laufzeitumgebung ist für das IT-Produkt notwendig? SUN mit Solaris 8/9, JAVA 1.4.2+, Apache 1.3+, Tomcat 4.18. Andere Umgebungen wären wohl möglich, sind aber nicht getestet. Bildserver: Python 2.3 mit PIL. Bisher wurde das IT-Produkt extern genutzt durch über 1100 registrierte Mitglieder bei der Wanderfalter-Beobachtung. Tägliche Dateneingabe. Jährlich kommen ca. 40.000 Meldungen und 5000 Belegfotos hinzu.

Geplante Erweiterungen: Mehrsprachigkeit, Ausbau der Kartendarstellung, Daten-Uploadtool für Excel-Daten, Audio- und Video-Upload.

## 35 SeSam

Datum der Ersteingabe: 21.04.2004.

Name der Institution: Forschungsinstitut und Naturmuseum Senckenberg, Frankfurt/Main.

Verantwortlicher für das IT-Produkt: Lothar Menner (Lothar.Menner@senckenberg.de).

Beteiligte Softwareentwickler: com2, Bad Homburg.

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: andere Lizenz (wird individuell entschieden).

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: ökologische Daten, Literaturverwaltung, wissenschaftliche Namen, Sammlungsbelege (Naturgeschichtliche Sammlungen), Bilddaten und geographische Daten.

Das IT-Produkt kann ausschließlich über das Internet genutzt werden.

Art des IT-Produktes: Datenbankanwendung und Anwenderprogramm.

Kurze Beschreibung: SeSam ist ein webbasiertes Sammlungsverwaltungsprogramm. Erfasst werden können zoologische und botanische Sammlungen (rezent und fossil). Viele Felder können den spezifischen Notwendigkeiten von Sammlungen angepasst werden. URL für das IT-Produkt: [sesam.senckenberg.de](http://sesam.senckenberg.de) (User: sesam).

Das IT-Produkt ist nicht modularisiert.

Für IT-Produkt mögliche HTTP Server: Microsoft IIS.

Das IT-Produkt läuft als stabile Version (1.0). Die in der Benutzeroberfläche verwendete Sprache ist deutsch und englisch. Ein Hilfesystem oder eine Benutzeranleitung ist in Bearbeitung. Eine technische Dokumentation ist in Bearbeitung. Programmiersprache(n) des IT-Produkts: andere. Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar (wird jedoch individuell entschieden). Die Quelltexte sind in deutscher Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: MS SQL Server (7). Ein publiziertes Datenmodell existiert nicht.

Datenbank Import-Format: Feldbegrenzer-limitierter Text. Datenbank Export-Format: Feldbegrenzer-limitierter Text. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien.

"Authority files" wurden integriert. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist nicht eingeschränkt. Die Datenbankinhalte können online global editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC (zwischen Datenbank und Webserver). Realisierung (Technologie) der Web Schnittstelle: ASP. Programmiersprache für Datenbank-Web Schnittstelle: VBScript und JavaScript.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Voraussetzung ist lediglich ein Webbrowser unabhängig vom Betriebssystem. Welche Laufzeitumgebung ist für das IT-Produkt notwendig? keine. Bisher wurde das IT-Produkt extern genutzt durch an GBIF beteiligte Institutionen im Knoten Evertebraten 3.

## 36 Specify

Datum der Ersteingabe: 21.04.2004.

Revisionsdatum: 4.7.2006.

Name der Institution: Specify Software Project, University of Kansas Biodiversity Research Center (Betreuung in DE: BGBM Berlin-Dahlem).

Verantwortlicher für das IT-Produkt: James Beach (jbeach@ku.edu).

Beteiligte Softwareentwickler: s. [http://www.specifysoftware.org/Specify/who\\_are\\_we/](http://www.specifysoftware.org/Specify/who_are_we/).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: OSI-zertifizierte open source Software. Bezeichnung der Lizenz: GPL.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei The Center for Research, University of Kansas, Lawrence, Kansas, USA.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Sammlungsbelege (Naturgeschichtliche Sammlungen), Beobachtungsdaten (Monitoring, Erfassungen), Bilddaten, geographische Daten, Schutzstatus, beschreibende Daten, Literaturverwaltung, Linksammlung und andere.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden (keine Online-Nutzung über das WWW, nur in lokalen Netzwerken) oder lokal genutzt werden.

Art des IT-Produktes: Datenbankanwendung, Anwenderprogramm, Web-Schnittstelle und Webservice, Wrapper (DIGIR, BIOCASE Wrapper verfügbar).

Kurze Beschreibung: Specify ist eine hochkonfigurierbare Datenbankanwendung für Microsoft Windows (2000, XP) auf SQL-Server/MSDE-Basis. Das Programm dient der Verwaltung und Pflege diverser naturhistorischer Sammlungen und unterstützt dabei zahlreiche Disziplinen der Botanik und Zoologie. Das Datenmodell von Specify ist hochkomplex und darauf ausgelegt, möglichst alle denkbaren Vorgaben und Ansprüche von Sammlungshaltern zu berücksichtigen. Zu diesem Zweck wurden mehr als 70 Tabellen mit über 800 verschiedenen Datenfeldern entwickelt und verknüpft. Specify wurde in jahrelanger Arbeit am Informatics Biodiversity Research Center der University of Kansas (USA) entwickelt und stetig verbessert. Das Programm steht auf der Website des amerikanischen Projektteams zum kostenlosen Download bereit: <http://www.specifysoftware.org>. URL für das IT-Produkt: <http://www.specifysoftware.org/>.

Das IT-Produkt ist modularisiert (Specify beinhaltet den RAVE Report Designer von Nevrona zum Gestalten von Etiketten und Berichten. Außerdem ist die Nutzung des Programms insofern modular, als das in viele Tabellen Daten eingegeben werden können, die nicht sofort mit Sammlungsobjekten oder Taxa verknüpft werden müssen (z. B. Literatur, Fundorte, Projekte etc.)).

Für IT-Produkt mögliche HTTP Server: Apache und Microsoft IIS.

Das IT-Produkt läuft als stabile Version. Die in der Benutzeroberfläche verwendete Sprache ist mehrsprachig (Momentan sind Englisch und Deutsch implementiert, weitere Sprachen sind vorgesehen und können sobald eine Übersetzung vorliegt implementiert werden.). Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar und offline verfügbar. Eine technische Dokumentation ist online verfügbar (vereinfachte Datenmodelle und ein Data Dictionary sind online, weitere technische Dokumentation ist z. Zt. nicht verfügbar, jedoch in Arbeit.) oder in Bearbeitung. Programmiersprache(n) des IT-Produkts: Delphi, VB und andere.

Entwicklungsumgebung(en): Windows 2000, XP.

Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar. Die Quelltexte sind z. T. (nur neuerer Code) in englischer Sprache kommentiert. Quelltext (URL): Anfrage richten an [jbeach@ku.edu](mailto:jbeach@ku.edu).

Verwendetes Datenbank-Verwaltungssystem: MS SQL Server. Ein publiziertes Datenmodell existiert (vereinfachtes Datenmodell ist online, komplexeres ist in Arbeit. Bislang wird alte MS Access Version des Datenmodells genutzt) oder ist in Bearbeitung.

Datenbank Export-Format: Feldbegrenzer-limitierter Text und andere (MS Access). Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien.

"Authority files" wurden integriert. Welche "authority files" können verwendet werden? Taxonomische Dateien im ITIS-Format. Eine Datenbankabfrage online ist mit statischer Dokumentdarstellung möglich oder mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist nicht eingeschränkt (Wird durch Datenhalter bestimmt). Die Datenbankinhalte können online nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC. Realisierung (Technologie) der Web Schnittstelle: CGI. Programmiersprache für Datenbank-Web Schnittstelle: andere (Delphi).

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Windows 2000, XP oder Windows Server 2003 für Specify / IIS oder Apache. Keine Client Programme. Verwaltung des Web Service über: UDDI (GBIF UDDI Registry). Interface für Wrapper: DiGIR oder BioCASE. Bisher wurde das IT-Produkt extern genutzt durch Weltweit über 100 registrierte Institutionen; in Deutschland bislang 6 Datenbanken, die durch das BGBM betreut werden und schon implementiert sind, weitere in Planung bzw. "selbständig" tätig.

### 37 SPICE-Wrapper

Datum der Ersteinreichung: 19.07.2007.

Name der Institution: Staatliche Naturwissenschaftliche Sammlungen Bayerns, IT-Zentrum.

Verantwortlicher für das IT-Produkt: Dr. Dagmar Triebel (triebel@bsm.mwn.de).

Beteiligte Softwareentwickler: Dieter Neubacher (neubacher@bsm.mwn.de).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: andere Lizenz. Bezeichnung der Lizenz: GPL.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen.

Das IT-Produkt kann ausschließlich über das Internet genutzt werden (Webservice).

Art des IT-Produktes: Datenbankanwendung und Webservice, Wrapper.

Kurze Beschreibung: stellt taxonomische Namen mit einiger Zusatzinformation für die Datenbank von Species 2000 als Webservice/ Wrapper zur Verfügung. URL für das IT-Produkt: <http://www.snsb.info/Wrappers/Spice/>.

Das IT-Produkt ist modularisiert (greift auf DiversityTaxonNames-Datenbank zu) oder nicht modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist nicht verfügbar. Eine technische Dokumentation ist nicht verfügbar. Programmiersprache(n) des IT-Produktes: Java. Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar (auf Anfrage). Die Quelltexte sind teilweise in englischer Sprache kommentiert.

Ein publiziertes Datenmodell existiert (publiziert durch Species2000). Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich (keine Browserschnittstelle). Die Datenbankabfrage online ist nicht eingeschränkt. Die Datenbankinhalte können online nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): JDBC. Programmiersprache für Datenbank-Web Schnittstelle: Java.

Verbindungsprotokoll für Webservices: SOAP. Bisher wurde das IT-Produkt extern genutzt durch Species 2000 Consortium.

Geplante Erweiterungen: Ausbau, Adaption an die Bedürfnisse von GBIF (im Rahmen des GBIF-Projektes Melastomataceae), Verbesserung der Dokumentation.

### 38 SYNTHESYS Data Quality Tool

Datum der Ersteinreichung: 9.7.2007.

Name der Institution: BGBM.

Verantwortlicher für das IT-Produkt: Anton Güntsch (a.guentsch@bgbm.org).

Beteiligte Softwareentwickler: Cristian Oancea (c.oancea@bgbm.org).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: OSI-zertifizierte open source Software. Bezeichnung der Lizenz: Mozilla Public Licence.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei BGBM.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: Sammlungsbelege (Naturgeschichtliche Sammlungen) und Beobachtungsdaten (Monitoring, Erfassungen).

Das IT-Produkt kann ausschließlich über das Internet genutzt werden.

Art des IT-Produktes: Webservice, Wrapper.

Kurze Beschreibung: Der Webservice dient der Kontrolle der Datenqualität von GBIF Datenanbietern. Für eine gegebene Provider-URL wird ein Report erstellt, der auf Probleme in den Daten hinweist. URL für das IT-Produkt: noch nicht veröffentlicht (geplant für 8/2007).

Das IT-Produkt ist nicht modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt in Entwicklung. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist in Bearbeitung. Eine technische Dokumentation ist in Bearbeitung. Programmiersprache(n) des IT-Produkts: Java.

Entwicklungsumgebung(en): Eclipse.

Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar. Die Quelltexte sind in englischer Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: MySQL. Realisierung (Technologie) der Web Schnittstelle: Java Servlets.

Verbindungsprotokoll für Webservices: andere. Bisher wurde das IT-Produkt extern genutzt. Der Webservice ist noch nicht veröffentlicht.

### **39 SysMyr (Version 4)**

Datum der Ersteintragung: 22.9.2004.

Revisionsdatum: 12.07.2007.

Name der Institution: Zoologische Staatssammlung München (ZSM).

Verantwortlicher für das IT-Produkt: Jörg Spelda.

Beteiligte Softwareentwickler: Jörg Spelda (spelda@zsm.mwn.de).

Die IT-Entwicklung ist für Dritte nicht verfügbar. Lizenztyp: Produkt für andere Institutionen nicht verfügbar.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei Jörg Spelda.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Synonyme, Sammlungsbelege (Naturgeschichtliche Sammlungen), Bilddaten, geographische Daten und andere.

Das IT-Produkt kann lokal genutzt werden oder über Client-Server Anbindung genutzt werden (auch über WWW abfragbar).

Art des IT-Produktes: Datenbankanwendung.

Kurze Beschreibung: Sysmyr3 erfasst vor allem Zitationen: Beschreibungen von Gattungen und Arten, Nachweise derselben (die auch georeferenziert werden können), Abbildungen, Synonymien (aus der Literatur). Daraus baut sich halbautomatisch der "taxonomic backbone" auf. URL für das IT-Produkt: <http://www.gbifev2.mwn.de>.

Das IT-Produkt ist modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt läuft als stabile Version. Die in der Benutzeroberfläche verwendete Sprache ist mehrsprachig (deutsch und englisch). Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar (für die Webabfrage) und nicht verfügbar. Eine technische Dokumentation ist nicht verfügbar. Programmiersprache(n) des IT-Produkts: VB, andere und C/C++ (Webservice in PHP programmiert).

Entwicklungsumgebung(en): MS Access 200, XAMPP.

Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar. Die Quelltexte sind nicht kommentiert.

Verwendetes Datenbank-Verwaltungssystem: MySQL, MS Access oder Oracle (entwickelt in MS Access, Webservice über MySQL). Ein publiziertes Datenmodell existiert nicht.

Datenbank Import-Format: Feldbegrenzer-limitierter Text. Datenbank Export-Format: Feldbegrenzer-limitierter Text. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien oder ODBC.

"Authority files" wurden nicht integriert. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist eingeschränkt. Die Datenbankinhalte können online nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnitt-

stelle): JSP. Realisierung (Technologie) der Web Schnittstelle: andere und CGI (XAMPP). Programmiersprache für Datenbank-Web Schnittstelle: PHP.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? WIN32 (ME, XP, 2000 getestet). Bisher wurde das IT-Produkt extern genutzt durch Species 2000 (Kooperationsvertrag besteht).

Geplante Erweiterungen: Weitere Datenfelder der SysMyr-Datenbank abfragbar zu machen. Derzeit sind gültige Namen, Synonyme und Literatur abfragbar. Als nächster Schritt sind geographische Daten geplant.

## 40 SysTax

Datum der Ersteingabe: 15.04.2004.

Revisionsdatum: 01.05.2007.

Name der Institution: Institut für Systematische Botanik und Ökologie, Universität Ulm.

Verantwortlicher für das IT-Produkt: Jürgen Hoppe (juergen.hoppe@uni-ulm.de).

Beteiligte Softwareentwickler: Evelin Boos (evelin.boos@snafu.de), Thorsten Ludwig (thorsten.ludwig@uni-ulm.de), Michael Wiedemann (michael.wiedemann@uni-ulm.de).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: andere Lizenz. Bezeichnung der Lizenz: Oracle Lizenz.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Synonyme, Sammlungsbelege (Naturgeschichtliche Sammlungen), Beobachtungsdaten (Monitoring, Erfassungen), Bilddaten, Populationsdaten, ökologische Daten, Schutzstatus, beschreibende Daten, Daten zur Präparation und Aufbewahrung und Literaturverwaltung (Pflanze-Tier-Interaktionen, Multimedia-Daten allgemein, Leihverkehr (Sammlungsverwaltung), Verwaltung Botanischer Gärten incl. Erstellung des Index Seminum)).

Das IT-Produkt kann über Client-Server Anbindung genutzt werden, lokal genutzt werden (SysTax-light als stand-alone-Version mit Export-Schnittstelle) oder über das Internet.

Art des IT-Produktes: Datenbankanwendung, Client-Software, Anwenderprogramm, Web-Schnittstelle und Webservice, Wrapper.

Kurze Beschreibung: SysTax ist ein multi-user, multi-institutionelles, konzeptbasiertes Datenbanksystem für Zoologie und Botanik. URL für das IT-Produkt: <http://www.biologie.uni-ulm.de/systax>.

Das IT-Produkt ist nicht modularisiert (Client besteht aus einzelnen Modulen im Sinne von ORACLE Forms).

Für IT-Produkt mögliche HTTP Server: Apache.

Das IT-Produkt läuft als Beta Testversion (4.1). Die in der Benutzeroberfläche verwendete Sprache ist mehrsprachig (deutsch und englisch). Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar. Eine technische Dokumentation ist online verfügbar. Programmiersprache(n) des IT-Produkts: andere (Oracle-Forms, Oracle-PL/SQL, Perl, C).

Entwicklungsumgebung(en): Oracle-Forms, vi.

Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar. Die Quelltexte sind in deutscher Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: Oracle (8.1.7.4). Ein publiziertes Datenmodell existiert in Form des ER-Diagramms und der Tabellenstruktur.

Datenbank Import-Format: XML und Feldbegrenzer-limitierter Text. Datenbank Export-Format: XML und Feldbegrenzer-limitierter Text. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien.

Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist nicht eingeschränkt (darüberhinaus besteht die Möglichkeit, sensible Daten vor einem Zugriff zu schützen). Die Datenbankinhalte können online nur von bestimmten Usern editiert werden und nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung,



Schnittstelle): andere. Realisierung (Technologie) der Web Schnittstelle: CGI. Programmiersprache für Datenbank-Web Schnittstelle: Perl und C/C++.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Windows 98, Windows NT, Windows 2000, Windows XP. Welche Laufzeitumgebung ist für das IT-Produkt notwendig? Oracle-Forms-Client 6i, wird mitgeliefert. Interface für Wrapper: BioCASE. Bisher wurde das IT-Produkt extern genutzt durch siehe: <http://www.biologie.uni-ulm.de/syntax/projects/partners.html> und <http://www.biologie.uni-ulm.de/syntax/infgard/summe.html>.

#### **41 TaxDB**

Datum der Ersteingabe: 28.9.04.

Name der Institution: FIS, Abteilung Deutsches Zentrum für Marine Biodiversitätsforschung (DZMB).

Verantwortlicher für das IT-Produkt: Burkhard Küster ([bkuester@senckenberg.de](mailto:bkuester@senckenberg.de)).

Die IT-Entwicklung ist für Dritte verfügbar (auf Anfrage).

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Synonyme, beschreibende Daten und Literaturverwaltung.

Art des IT-Produktes: Datenbankanwendung.

Das IT-Produkt ist in Planung.

#### **42 Taxonomischer Webeditor für Berlin Modell Datenbanken**

Datum der Ersteingabe: 15.04.2004.

Revisionsdatum: 30.6.2006.

Name der Institution: Botanischer Garten und Botanisches Museum Berlin-Dahlem, Abt. Biodiversität-sinformatik.

Verantwortlicher für das IT-Produkt: W.Berendsohn ([w.berendsohn@bgbm.org](mailto:w.berendsohn@bgbm.org)), A.Güntsch ([a.guentsch@bgbm.org](mailto:a.guentsch@bgbm.org)). Beteiligte Softwareentwickler: A. Güntsch ([a.guentsch@bgbm.org](mailto:a.guentsch@bgbm.org)), P. Ciardelli ([p.ciardelli@bgbm.org](mailto:p.ciardelli@bgbm.org)), M. Geoffroy ([m.geoffroy@bgbm.org](mailto:m.geoffroy@bgbm.org)).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: andere Lizenz. Bezeichnung der Lizenz: Mozilla Public Licence.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei BGBM.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen (Namen und Potential Taxa), Trivialnamen und Bilddaten.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden.

Art des IT-Produktes: Datenbankanwendung, Client-Software, Anwenderprogramm und Web-Schnittstelle.

Kurze Beschreibung: Webeditor für Taxonomische Datenbanken, die das Berlin-Modell implementieren. Die Software basiert auf dem "Potential Taxon" Konzept und ist in der Lage konkurrierende taxonomische Auffassungen für taxonomische Gruppen zu verarbeiten. URL für das IT-Produkt: <http://www.bgbm.org/biodivinf/docs/bgbm-model/>.

Das IT-Produkt ist modularisiert.

Für IT-Produkt mögliche HTTP Server: Apache und Microsoft IIS.

Das IT-Produkt läuft als Beta Testversion. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar und offline verfügbar. Eine technische Dokumentation ist offline verfügbar. Programmiersprache(n) des IT-Produktes: Java und andere (ColdFusion MX).

Entwicklungsumgebung(en): Eclipse, Macromedia Dreamweaver.

Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar. Die Quelltexte sind in englischer Sprache kommentiert. Quelltext (URL): <http://ww2.biocase.org/websvn/>.

Verwendetes Datenbank-Verwaltungssystem: MS SQL Server. Ein publiziertes Datenmodell existiert. Bezeichnung des Datenmodells: Berlin Modell (s. <http://www.bgbm.org/biodivinf/docs/bgbm-model/>).

Datenbank Import-Format: XML und andere (Datenimporte jetzt auch über spezielles Word-Format möglich.). Datenbank Export-Format: Feldebegrenzer-limitierter Text. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien.

Eine Datenbankabfrage online ist mit statischer Dokumentdarstellung möglich. Die Datenbankabfrage online ist nicht eingeschränkt. Die Datenbankinhalte können online nur von bestimmten Usern editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC oder JDBC. Realisierung (Technologie) der Web Schnittstelle: ASP.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Läuft auf allen Betriebssystemen (nur Javascript fähiger Browser benötigt), auf dem Server wird ColdFusion MX benötigt (läuft auf Windows32 und Linux). Welche Laufzeitumgebung ist für das IT-Produkt notwendig? Java 1.4. Interface für Wrapper: BioCASE (Berlin-Modell datenbanken sind über BioCASE Wrapper an Species2000 angeschlossen). Bisher wurde das IT-Produkt extern genutzt durch Bundesamt für Naturschutz, Euro+Med Sekreatariat (School of Plant Sciences in Reading), Institute of Botany, Academy of Sciences of the Czech Republic, Species2000.

### 43 Unitloader

Datum der Ersteingabe: 15.04.2004.

Revisionsdatum: 30.6.2006.

Name der Institution: Botanischer Garten und Botanisches Museum Berlin-Dahlem, Abt. Biodiversität-sinformatik.

Verantwortlicher für das IT-Produkt: W.Berendsohn ([w.berendsohn@bgbm.org](mailto:w.berendsohn@bgbm.org)), A.Güntsch ([a.guentsch@bgbm.org](mailto:a.guentsch@bgbm.org)).

Beteiligte Softwareentwickler: A.Güntsch ([a.guentsch@bgbm.org](mailto:a.guentsch@bgbm.org)).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: OSI-zertifizierte open source Software. Bezeichnung der Lizenz: Mozilla Public Licence.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei BGBM.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: Synonyme und Sammlungsbelege (Naturgeschichtliche Sammlungen).

Das IT-Produkt kann lokal genutzt werden.

Art des IT-Produktes: andere (Java API).

Kurze Beschreibung: Java API zur Erzeugung und parallelen Versendung von Anfragen an verteilte BioCASE Datenanbieter. URL für das IT-Produkt: <http://www.biocase.org/dev/unitloader/>.

Das IT-Produkt ist modularisiert, läuft als Beta Testversion. Ein Hilfesystem oder eine Benutzeranleitung ist online verfügbar und offline verfügbar. Eine technische Dokumentation ist online verfügbar oder offline verfügbar. Programmiersprache(n) des IT-Produkts: Java.

Entwicklungsumgebung(en): Borland JBuilder.

Die Quelltexte der IT-Entwicklung sind für Dritte verfügbar. Die Quelltexte sind in englischer Sprache kommentiert. Quelltext (URL): <http://www.biocase.org/dev/unitloader/>. Quelltext (CVS Account): cvs nicht öffentlich.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Getestet auf Windows NT, 2000, 98, Linux. Welche Laufzeitumgebung ist für das IT-Produkt notwendig? Java 1.4. Bisher wurde das IT-Produkt extern genutzt durch UPMC/LIS – Laboratoire Informatique et Systématique, Université Paris 6.

Der UnitLoader soll nicht weiterentwickelt werden.

## 44 VegDa

Datum der Ersteingabe: 10.07.2007.

Name der Institution: Forschungsinstitut Senckenberg, J.W.Goethe-Universität Frankfurt.

Verantwortlicher für das IT-Produkt: Marco Schmidt (mschmidt@senckenberg.de).

Die IT-Entwicklung ist für Dritte verfügbar (auf Anfrage). Lizenztyp: Lizenzpolitik ungeklärt.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei Marco Schmidt.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: Beobachtungsdaten (Monitoring, Erfassungen) (Vegetationsaufnahmen nach verschiedenen Methoden).

Das IT-Produkt kann lokal genutzt werden.

Art des IT-Produktes: Datenbankanwendung (auf Basis von MS Access).

Kurze Beschreibung: VegDa ist eine Datenbank für Vegetationsaufnahmen auf der Basis von Microsoft Access. Es können Aufnahmen nach Braun-Blanquet, nach BIOTA-Konventionen, oder einfache Artenlisten verwaltet werden. Indices wie Shannon-Weaver, Simpson und Evenness werden von der Datenbank errechnet, über Export in eine csv-Datei sind weitere Analysen mit Programmen wie CAP, Canoco und SPSS möglich, auch für Programme zur Verbreitungsmodellierung wie GARP und MaxEnt besteht eine Exportfunktion.

Das IT-Produkt ist nicht modularisiert. Die in der Benutzeroberfläche verwendete Sprache ist englisch. Ein Hilfesystem oder eine Benutzeranleitung ist in Bearbeitung. Eine technische Dokumentation ist nicht verfügbar. Verwendetes Datenbank-Verwaltungssystem: MS Access. Ein publiziertes Datenmodell existiert nicht.

Datenbank Export-Format: Feldbegrenzer-limitierter Text.

Geplante Erweiterungen: XML Import/Export (ESVeg-kompatibel).

## 45 Visual Plants

Datum der Ersteingabe: 19.07.2007.

Name der Institution: Universität Hohenheim, Institut für Botanik.

Verantwortlicher für das IT-Produkt: Helmut Dalitz (hdalitz@uni-hohenheim.de).

Beteiligte Softwareentwickler: Jörg Waltemathe (jwaltema@uni-hohenheim.de), Liat Mizrahi-Todt (mizrahil@uni-hohenheim.de), Dr. Jürgen Homeier (jhomeie@gwdg.de).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: andere Lizenz. Bezeichnung der Lizenz: Über die Dalitz & Homeier GbR beziehbar, <http://www.visualplants.de>.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei der Dalitz & Homeier GbR.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen (Pflanzenmerkmale; über Multi-access key suchbar), Trivialnamen, Synonyme, Beobachtungsdaten (Monitoring, Erfassungen), Bilddaten, geographische Daten und beschreibende Daten.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden oder lokal genutzt werden.

Art des IT-Produktes: Datenbankanwendung, Web-Schnittstelle und Webservice, Wrapper.

Kurze Beschreibung: Visual Plants wurde in 2000 begonnen als reine Verwaltung von Bilddaten von Pflanzen. Durch die Mitwirkung in zwei großen Forschungsverbänden (und damit einer Finanzierung) konnte die Weiterentwicklung zu einer Web-basierten Anwendung (neben der bestehenden lokalen) betrieben werden. Die Aufnahme von Voucher-Specimen mit Bildmaterial stand dabei im Fordergrund. In 2005 wurde die Entwicklung einer Volltextsuche in beschreibenden Texten begonnen. URL für das IT-Produkt: <http://www.visualplants.de>.

Das IT-Produkt ist nicht modularisiert.

Für IT-Produkt mögliche HTTP Server: andere und Apache (Als Webserver wird der im RAD-Tool 4D angebotene Webserver verwendet.).

Das IT-Produkt läuft als Beta Testversion oder ist erst in Planung (2.16). Die in der Benutzeroberfläche verwendete Sprache ist englisch, andere und deutsch (mehrsprachiges Interface in Vorbereitung). Ein Hilfesystem oder eine Benutzeranleitung ist nicht verfügbar. Eine technische Dokumentation ist in Bearbeitung. Programmiersprache(n) des IT-Produkts: C/C++ (Es wird die in dem RAD-Tool angebotene Programmiersprache (ähnlich C++) verwendet.).

Entwicklungsumgebung(en): 4D (<http://www.4d.com>), kommerzielles Produkt.

Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar. Die Quelltexte sind in englischer Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: andere oder Oracle (4D (<http://www.4d.com>)). Ein Datenmodell zur Publikation ist in Bearbeitung.

Datenbank Import-Format: XML und Feldbegrenzer-limitierter Text. Datenbank Export-Format: XML und Feldbegrenzer-limitierter Text. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien (bisher nur über Import/Export; ODBC für Datenaustausch mit GBIF).

"Authority files" wurden nicht integriert. Eine Datenbankabfrage online ist mit dynamischer Dokumentengenerierung möglich. Die Datenbankabfrage online ist eingeschränkt (Lokalitätsinformationen sind auf Wunsch der Projektpartner nur für registrierte Nutzer möglich. Registrierung erfordert Email-Kontakt unter Beschreibung des Nutzungszwecks.). Die Datenbankinhalte können online noch nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC (für GBIF über ODBC, sonst proprietär) oder andere. Realisierung (Technologie) der Web Schnittstelle: CGI (Nutzung der Webschnittstelle von 4D) und andere. Programmiersprache für Datenbank-Web Schnittstelle: andere und PHP (Nutzung der Programmierumgebung von 4D).

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Windows XP oder Mac OSX; Download des 4D Clients (<http://www.4d.com>); 4 parallele Nutzungen sind derzeit möglich.

Die lokale Version ist compiliert und uneingeschränkt nutzbar für die oben angegebenen Betriebssysteme. Welche Laufzeitumgebung ist für das IT-Produkt notwendig? s.o. Interface für Wrapper: BioCASE. Bisher wurde das IT-Produkt extern genutzt durch National Museum of Kenya, Botany Department, Nairobi, Kenya; Herbarium of the Makerere University, Kampala, Uganda; Universidad de Costa Rica, Sede de Occidente, San Ramon, Costa Rica Universidad Technica Particular de Loja, Loja, Ecuador;.

Geplante Erweiterungen: Komplette Umsetzung des ABCD-Standards, Verbesserung des Web-Interfaces, Web-Basierte Dateneingabe; Nutzung von Authority files (Geographical names; accepted names for organisms; Synonyms, local names).

## 46 WISIA

Datum der Ersteingabe: 27.9.2004.

Revisionsdatum: 27.04.2007.

Name der Institution: Bundesamt für Naturschutz.

Verantwortlicher für das IT-Produkt: Iris Göde.

Beteiligte Softwareentwickler: science&communication Bonn, Norbert Hirneisen (nobby@s2you.de).

Die IT-Entwicklung ist für Dritte verfügbar. Lizenztyp: Lizenzpolitik ungeklärt oder OSI-zertifizierte open source Software (Einzelfallklärung).

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Synonyme, geographische Daten, Schutzstatus, beschreibende Daten und Literaturverwaltung.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden.

Art des IT-Produktes: Datenbankanwendung, Client-Software und Anwenderprogramm.

Das IT-Produkt ist nicht modularisiert, läuft als stabile Version. Die in der Benutzeroberfläche verwendete Sprache ist deutsch. Ein Hilfesystem oder eine Benutzeranleitung ist offline verfügbar. Eine tech-

nische Dokumentation ist offline verfügbar. Programmiersprache(n) des IT-Produkts: C/C++, Delphi und andere.

Entwicklungsumgebung(en): Borland IDE.

Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar. Die Quelltexte sind in deutscher Sprache kommentiert.

Verwendetes Datenbank-Verwaltungssystem: Oracle. Ein publiziertes Datenmodell existiert nicht.

Datenbank Import-Format: Feldbegrenzer-limitierter Text. Datenbank Export-Format: Feldbegrenzer-limitierter Text. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien.

"Authority files" wurden integriert. Welche "authority files" können verwendet werden? Taxonomische Referenz. Eine Datenbankabfrage online ist nicht möglich.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Windows XP/2000.

## **47 Zoologische und botanische Funddaten**

Datum der Ersteingabe: 21.4.04.

Revisionsdatum: 22.4.07.

Name der Institution: Nationalpark Berchtesgaden.

Verantwortlicher für das IT-Produkt: Helmut Franz (H.Franz@nationalpark-berchtesgaden.de).

Beteiligte Softwareentwickler: Doris Huber (D.Huber@nationalpark-berchtesgaden.de) (Botanik), Annette Lotz (A.Lotz@nationalpark-berchtesgaden.de) (Zoologie).

Die IT-Entwicklung ist für Dritte nicht verfügbar. Lizenztyp: Produkt für andere Institutionen nicht verfügbar.

Copyright für Eigenentwicklungen am Softwareprodukt liegt bei Nationalpark Berchtesgaden.

Art der verwalteten bzw. zu analysierenden Biodiversitätsdaten: wissenschaftliche Namen, Trivialnamen, Sammlungsbelege (Naturgeschichtliche Sammlungen), Bilddaten, Audio-Daten, geographische Daten, molekulare Daten und Populationsdaten.

Das IT-Produkt kann über Client-Server Anbindung genutzt werden.

Art des IT-Produktes: Datenbankanwendung (und GIS-Anwendung).

Kurze Beschreibung: Zoologische und botanische Funddaten werden nach einer von der Nationalparkverwaltung vorgegebenen Struktur digital im Rahmen der Forschungsprojekte aufbereitet und in die zentrale Datenbank der Nationalparkverwaltung überführt. Es werden die Datenbanksoftware Oracle und die GIS-Software ArcGis eingesetzt.

Das IT-Produkt ist modularisiert, in Entwicklung. Die in der Benutzeroberfläche verwendete Sprache ist deutsch. Ein Hilfesystem oder eine Benutzeranleitung ist nicht verfügbar. Eine technische Dokumentation ist nicht verfügbar. Programmiersprache(n) des IT-Produkts: andere (Oracle-Tools). Die Quelltexte der IT-Entwicklung sind für Dritte nicht verfügbar.

Sind die Quelltexte kommentiert? Nein.

Verwendetes Datenbank-Verwaltungssystem: Oracle (Version 10g + ArcGIS 9.2). Ein publiziertes Datenmodell existiert. Bezeichnung des Datenmodells: ZOOLIS, BOTIS.

Datenbank Import-Format: Feldbegrenzer-limitierter Text. Datenbank Export-Format: Feldbegrenzer-limitierter Text. Datenbank-Datenbank-Content-Transfer über: Import/Export-Dateien oder ODBC.

"Authority files" wurden nicht integriert. Eine Datenbankabfrage online ist nicht möglich. Die Datenbankabfrage online ist nicht eingeschränkt. Die Datenbankinhalte können online nicht editiert werden. Zugriff auf die Datenbank über (Web-Datenbankanbindung, Schnittstelle): ODBC, JDBC oder JSP.

Welches Betriebssystem ist für Client bzw. Anwendungsprogramm notwendig? Windows XP. Bisher wurde das IT-Produkt nicht extern genutzt.

Geplante Änderungen: Portierung auf des RZ Süd des Freistaats Bayern, Anschluss über Oracle Tools (http) und ArcIMS für Internet-Anwendungen.

