

„Portalentwicklung im GBIF-Verbund“

Protokoll zum Treffen der GBIF-D IT-Kommission
am 18.01.2012 von 11.00 – 16.00 Uhr



Tagungsort: Senckenbergmuseum, Frankfurt

Teilnehmer:

Wolfram Fritzsch (MfN), Maren Gleisberg (BGBM), Anton Güntsch (BGBM), Gregor Hagedorn (JKI), Jörg Holetschek (BGBM), Alexander Kroupa (MfN), Michael Martin (DSMZ), Lothar Menner (Senckenberg), Martin Müller (Senckenberg), Dagmar Triebel (SNSB), Michael Tuerkay (Senckenberg; bis 13.00Uhr), Tanja Weibulat (SNSB)

Entschuldigt: Peter Grobe (ZFMK)

Michael Tuerkay begrüßt die IT Kommission am Hause Senckenberg. Anton Güntsch begrüßt die anwesenden Kommissionsmitglieder und Gäste. Die Tagesordnung wird unverändert angenommen.

TOP 1: Nachlese letzte Kommissionssitzung

- **Bildserver: Stand der Kooperationen [Güntsch]**

Eine beispielhafte Verarbeitung der Bilder auf dem BGBM Bildserver wurde exemplarisch mit dem DSMZ durchgeführt. Hierfür wurde ein Nutzerkonto für das DSMZ auf dem Bildserver eingerichtet. Erste Bilddaten sind hochgeladen und erfolgreich getestet. (Beispiel von Herr Martin). In einem nächsten Schritt wird nun das BGBM Programmcode liefern, mit dem das DSMZ Bilder in eigene Webpräsentationen einbinden kann.

Eine ähnliche Kooperation mit der Uni Heidelberg läuft bereits. Die Services sind offen für alle Projektpartner, bei Interesse mit dem BGBM in Kontakt treten.

Rückfrage M. Martin/DSMZ Satz aus letztem Protokoll, 23.09.2011: „Anton Güntsch weist darauf hin, dass der BGBM zwar in Einzelfällen das Hosting von Bilddatenbeständen übernehmen kann, dies aber als generellen Service aus Kapazitätsgründen nicht anbieten kann.“. Dazu Anton Güntsch: Das BGBM bietet als Service an, Bilddaten im Kontext von GBIF-D im Cache des Imageservers zu speichern und im Web zu publizieren. Die Originaldaten sollen aber in der Regel beim Urheber gespeichert und gepflegt werden. In Ausnahmefällen kann das BGBM auch Originaldaten hosten, dieses aber (mangels Kapazität) nicht als generellen Service anbieten

Das jeweilige Institut kümmert sich darum, dass Bilddaten gespeichert und gepflegt sind. Bei höheren Datenvolumina können die Bilddaten auch über Festplatten an das BGBM geschickt werden. Erfahrungswert von DSMZ: z. Zt. 90 Bilder gescannt, (Kalkulation: Hiwi braucht ca. 8 Monate f. 1000 Bilder, inkl. Metadaten). Die genauen Verfahrensweisen werden sich aus den Pilotanwendungen ergeben.

Mit Bilddaten aus Senckenberg ist zu rechnen, aber noch Vorarbeiten/Kommunikation hausintern erforderlich. Herbarbögen Senckenberg liegen derzeit nicht bei Senckenberg selbst.

- **Vorstellung des Wikis [Hagedorn / Triebel]**

Aus dem Protokoll vom 23.09.2011: Digitalisierungstechniken und Workflows (hierzu wird von Triebel/Hagedorn ein Wiki eingerichtet). Ziel ist unter anderem ein Best Practice Handbook

Das WIKI ist eingerichtet und erreichbar unter: <http://wiki.gbif.de>, <http://wiki.gbif.de/web>

G. Hagedorn: Zugriffssteuerung ist z.Zt. noch nicht geregelt. Welche Teile wie geschützt werden lässt sich an dem Namen der Wiki-Unterseite dann erkennen. Namensräume sind zunächst zu klären.

Das Wiki wird zunächst als geschlossene Seite auf Deutsch betrieben, die die jeweiligen Digitalisierungsstationen der Institute beschreibt und Vergleiche ermöglicht. Später kann daraus dann möglicherweise eine Publikation (zum Beispiel als Best Practice Handbook) entstehen. *(Ergänzend aus der Diskussion zum Protokoll G.Hagedorn: „Ja, ich hatte gesagt dass wir über die genau Zugriffsgestaltung reden, sobald klare Vorstellungen vorliegen wer was schreiben will und wie jeder das schützen möchte.“)*

Beteiligen wollen sich: MfN, Senckenberg, SNSB, BGBM.

Struktur: an den Arbeitsstationen aufzuhängen, Gregor braucht von den Teilnehmenden Institutionen Titelvorschläge und kann gruppieren/vereinheitlichen und entsprechende Seiten einrichten.

Aufgabe: In 8 Tagen (Do.26.01.2012)Titelvorschläge der Wiki-Seiten an Gregor Hagedorn senden.

Realistischer Zeitplan: Spätestens zum nächsten Jahresbericht im April 2013 sind die Wiki-Seiten fertigzustellen und öffentlich zu machen. Ein bereits früheres Einstellen der Handhabung der eigenen Digitalisierungsstation als Hilfe und Anleitung für andere ist sinnvoll.

Daran anschließend ist eine Publikation möglich. Gegenwärtig gibt es im Übrigen ein „special volume of ZooKeys on mass digitization of natural history collections“. (Diese ist lediglich als Beispiel zu verstehen). Von BMC Bioinformatics wurde ebenfalls ein Special Issue zum Thema Massendigitalisierung herausgegeben: Data publishing framework for primary biodiversity data:

<http://www.biomedcentral.com/bmcbioinformatics/supplements/12/S15>

A.Kroupa: Schlägt einen Workshop, um weitere Methoden / Partner zu finden, vor. D.Triebel und A.Güntsch wenden ein, dass dies nicht Aufgabe der IT K. sei.

A.Kroupa: MfN hat Digitalisierungen im Projektantrag als WP, und bietet die Organisation eines Workshops dazu an. Hier kann dann auch das fertige GBIF-D-Digitalisierungs-Wiki vorgestellt werden.

- **Mailingliste** [Güntsch / Gleisberg]

gbif-d-it-kommission@lists.fu-berlin.de

Diese Mailing-Liste wird ab sofort von allen für Kommissions-interne Kommunikation genutzt. Maren Gleisberg wird die Anmeldeinfos und den Archivlink ins Wiki stellen.

Nachträglich: Tanja Weibulat in Verteiler aufgenommen.

TOP 2: Portalentwicklung im GBIF-Verbund (11:45-14.30 // MP 12.30/13)

- **Datenfluss in Biodiversitätsnetzwerken (u.a. BioCASE) [Holetschek, siehe ppt]**
 - Harvesting/ Indexing: Alle 1-2 Monate erstellt GBIF eine neue Indexdatenbank.
 - Neu in GBIF ist die Nutzung des DwC-Archive Formates zur Datenübertragung vom Provider zum Index.
 - Das DwC-A Format kann sowohl vom IPT als auch der neuen Version der BioCASE Provider-Software geliefert werden.

Für die Erstellung eigener Index-Datenbanken kann das GBIF HIT-Tool genutzt werden. Dieses wir zum Beispiel erfolgreich im OpenUp! Projekt eingesetzt.

D. Triebel: Vorschlag ein deutsches Portal, v.a. für die Zielgruppen Projektträger und deutsche „Amateure“ – Ziel wäre es die Daten mit Bezug Deutschland besser zu präsentieren und Lücken zu identifizieren.

A.Güntsch: Wenn das Portal keine Extra-Funktionalitäten beinhaltet, könnte es in 2-4 Wochen aufgesetzt werden. Vordefinierte Abfragen für spezielle taxonomische Gruppen wären nur mit zusätzlichem Aufwand zu realisieren.

Vorschlag M. Gleisberg: Übersicht der von den Knoten betriebenen Portale ließe sich auf www.gbif.de darstellen. Website Umgestaltung ist geplant für 2012. Denkbar ist eine Seite mit allen Portalen und den Bereich prominenter darstellen. In dem Zusammenhang der Navigationsüberarbeitung auch für den „Citizen Scientist“ leicht zugänglich zu machen, Zeitplanung zu Mitte Ende des Jahres realistisch.

Vorschlag A. Güntsch: BGBM macht eine Abschätzung für zwei Datenportalvarianten als Eingabe für die Diskussion in der Knotenkoordinatorengruppe.

Weitere Technische Aspekte / Fragen?

Die Tools von GBIF sind frei verfügbar, HIT funktioniert gut.

Frage L. Menner: Wie liefere ich eine DwC-A-Datei an GBIF?

J. Holetschek: Aktualisierte Datei muss lediglich auf dem Webserver abgespeichert und die URL an GBIF geliefert werden. Das Verfahren bietet sich vor allem für große Datenmengen an.

- **Vorstellung der geplanten Entwicklungen durch die Partner**

BGBM / Botanik A. Güntsch: Implementierung eines virt. Herbariums, vgl. pdf

Scope: Herkunft der Belege Deutschland, geogr. Bezug weltweit

Beispiel: Vorstellung der Suchfunktionalität (simple und advanced search) im Specimen Editor in EDIT unter <http://search.biocase.org/edit>; Vorteil Suchfunktionen und Schnelligkeit

Anregung Gregor: Die Existenz thematischer Portale sollte bei GBIF-International besser vermarktet werden. Eingabe für Diskussion in der Knotenkoordinatorengruppe.

Frage D. Triebel: Umfasst das Herbar auch Pilze, Antrag nichts zu Herbarentwicklung beantragt und derzeit keine Entwicklung geplant.

BGBM / BSM Abstimmungsbedarf für Pilze, zunächst wird BGBM erst einmal Botanik ohne Pilze machen.

DSMZ, M. Martin: vgl. pdf Server Structure

Daten zu Bakterien und Archaeen sind selten georeferenziert (Geografic Origin)

Kultivierungskonditionen, Exportbestimmungen als zusätzliche Parameter der DSMZ-Daten

DSMZ hat technisch stark nachgerüstet um mit den Vorhaben, die u.a. via GBIF beantragt sind, Digitalisierungen beginnen zu können.

Ziel: Erstbelege und Sequencer-Belege sollen frei verfügbar über ein Portal zur Verfügung gestellt werden. Bsp.: Freie Suchen wie „gib mir alle Ölabbauer“ sind im Detail noch zu klären. Ein eigenes Portal ist geplant, da GBIF Suche eine rein taxonomische Suche ist, die für diese Zwecke nicht ausreicht. Focus der DSMZ: Datenqualität.

Weiteres Thema: Daten aus Textdaten (mögl. Themenvorschlag für IT Kommission: effiziente Erschließung von Textdaten).

SNSB, D. Triebel:

An der SNSB ist keine eigene Portalentwicklung im Arbeitspaket vorgesehen, Fokus ist die Datenbank-Backbone Weiterentwicklung und der technische Service für andere Knoten.

SNSB betreibt mehrere Portale mit Bezug zu / Optimierung auf speziellen Sammlungen

Vgl. <http://www.gbif-mykologie.de/Datenbanken.html> sowie andere Portale, die Daten anbieten mit direktem Zugriff auf GBIF int. So z.B. GBIF-D Mykologie Datenportale unter http://www.gbif-mykologie.de/Daten_Andere.html. Und Portale für Spezialsammlungen: http://www.botanischestaatssammlung.de/DatabaseClients/BSMlichfungicoll/DiversityCollection_BSMlichfungicoll_Find.cfm

SNSB verzeichnet steigende Zugriffsraten auf den Seiten sowie steigende Ausleihen im konkreten Bezug auf die Portal-Nr. Weitere technische Entwicklungen sind derzeit nicht vorgesehen.

MfN / A. Kroupa (siehe pdf)

Verwendung des Cumulus Systems. Für GBIF und v.a. OpenUp/Europeana wäre ein Webclient wichtig. Grundsätzliche Entscheidungen zur Web-Freigabe von Bildmaterial steht noch aus (hausinterne Überzeugungsarbeit). Bilder in MfN sollen in MfN-Portal dargestellt werden, möglicherweise über Cumulus. Scan Operator ist seit Jan. 2012 eingestellt, sichtet und testet derzeit Technik.

Vorstellung eines externen Beispiels einer Wiki-basierten Lösung für Feldtagebücher (vergleichbares Citizen Science Projekt auch am MfN für Eingangstagebücher geplant).

Erfahrungsbeispiel: 3D Objekterfassung Erfassung testweise mit FU zusammen.

Beispiel Hummel. z. Zt. 1.500 hochauflösende Bilder um 1 Hummel „drehbar“ zu machen; 1Std. z Fotografieren / dann ca. 6.Std Bildberechnung. Derzeit Austestung wie viel technisch leistbar ist. Ca. 5-15 GB pro Objekt. (*FU TIT-Projekt, Physik, Kontakt M.Pluta*)

A.Güntsch: Auch für Botanik wären 3D-Belege denkbar (Bsp. Tannenzapfen).

Eine Größenbeschränkung gibt es lediglich nach unten, vrsl. 4-5 mm Mindestobjektgröße, ist noch in Erprobung. Zur Langen Nacht der Museen am 28.01.2012 stellt MfN den 3D Scan einer Mordwespe auf Großleinwand vor.

Senckenberg, L. Menner:

SeSam als Portal, Version 2.0 wird derzeit entworfen.

Es gibt eine GIS-Komponente im Antrag: Bei Erfassung und Suche sollen GIS Funktionalitäten genutzt werden. Erfassung: Koordinaten angeben und geogr. Indexierung erfolgt dann automatisch. Suche/Kartendarstellung: in Zusammenarbeit mit einer GIS Firma, Ziel sind open source projekte, vrsl. in 2 Monaten einsatzfähig.

L. Menner wird diese GIS Funktionalitäten in einer nä. Sitzung vorstellen.

Bis Ende Februar 2012 soll das Produkt einsatzfähig sein, Prototyp wird dann via Mailingliste an die anderen Knoten gemailt.

- **Diskussion eines möglichen gemeinsamen Vorgehens**

Fazit: Vorgehensweise in den Knoten sehr unterschiedlich. SNSB und BGBM haben zumindest partiell Abstimmungsbedarf. MfN und ZFMK zu 3D und Citizen Science Abstimmen A. Kroupa kontaktiert P. Grobe dazu direkt.

3 Dinge: GBIF Portal, Knoteninstitutsportale, deutsches GBIF-Portal

Zum GBIF.de-Webauftritt:

Highlights der GBIF-D Knoten sollen besser dargestellt werden. „Warum soll ich meine Daten an GBIF geben“ (Frage des Hobby Wissenschaftlers) auf neuer Website besser beantworten.

Anregung (an Knotenkoordinatoren): Den Navigationsbutton „Daten/Datenabfrage“ bei allen Knoten durchsetzen! Beim Websiteumbau den Bereich von 2 Seiten (innerhalb der Knoten und in übergeordneter Navigation) aus erreichbar machen.

Maren Gleisberg trägt diesen und den folgenden **Vorschlag IT Kommission im Knotentreffen vor**: Highlight-Seite von extern per Werkvertrag erstellen lassen oder von NeFo / S. Tilch draufschauen lassen.

TOP 3: Arbeitsauftrag aus Knotentreffen: Bilderverwaltung / Fotoarchiv (14.30 – 15.30)

A. Güntsch stellt den Arbeitsauftrag aus dem Knotentreffen in München, 29.9.2011 vor. Die IT Kommission wird gebeten eine technische Lösung für die Bereitstellung von Fotos für Zwecke der Öffentlichkeitsarbeit von GBIF-D zu finden.

Drupal CMS oder Wiki (G.Hagedorn):

Bio-Wiki-Farm (www.biowikifarm.net/meta/Biowikifarm): Special Media Lizenz, d.h. keine Lizenz (niemand darf sie einfach nutzen, nur mit schriftlicher Genehmigung). Open media (alle mögl Creative Commons Lizenzen verwendbar, sowohl open Content Lizenzen wie cc-by (open access, pensoft etc.) und cc-by-sa (wikipedia). Creative Commons: Closed content wie cc-by-nc-sa (vgl. Zookeys).

Empfehlung Gregor wäre Open Media Lizenz: 2 Repositorien erforderlich: 1 Lizenziert, Rest Spezialbilder

Empfehlung an Knotengremium: Biowikifarm und Open Media Lizenz, je Knoten zunächst 5-10 Bilder einbinden. Die Knoten sollen hierfür Bilder mit Lizenzbestimmungen an Maren Gleisberg schicken.

TOP 4: Neue Projekte und Anträge mit GBIF-Bezug (15.30 – 15.45)

DFG Datacenter (Güntsch/Triebel)

Ein Nationales Datenservice-Zentrum für DFG-Projekte soll entstehen: Datenrettung, Prozessierung, Bereitstellung. Derzeit Antragsphase der beteiligten Projektpartner: Pangea Wold Data Center für marine Daten, SNSB, BGBM, Uni Jena (Informatik), Uni Marbug (Informatik), FU Berlin (Informatik), ZFMK, SMNS, DSMZ, MPI für Marine Microbiologie Bremen. Es ist ein sehr breit angelegtes Datenspektrum.

TOP 5: Abstimmungen zur weiteren Arbeit der IT-Kommission, Thema, Termin und Ort für das nächste Treffen

Workflows: Taverna (EU-System) / Kepler (Amerikanisches System), z.B. Projekt BioVel
L. Menner wird die GIS Funktionalitäten vorstellen.

Das nächste Treffen (ein Tag, Zeitraum wieder 11.00-16.00) findet in Braunschweig im September (ersten zwei Septemberwochen) statt. Der BGBM wird einen Doodle-Poll dazu einrichten und Braunschweig bei der Organisation unterstützen. Wir haben Gelegenheit auch das Institut und die Sammlung zu besichtigen.