

Protokoll des Treffens der GBIF-D IT-Kommission
am 21.11.2013 von 11.30 – 16.00 Uhr

„Datenflüsse zu und in GBIF (international) und
Identifikation von Fehlentwicklungen“



Tagungsort: SNSB München - Menzinger Straße 67 - 80638 München, Raum: 139

Teilnehmer:

Maren Gleisberg (BGBM), Anton Güntsch (BGBM), Peter Grobe (ZFMK, ab TOP3), Gregor Hagedorn (MfN, ab TOP2), Jörg Holetschek (BGBM), Daniel Mosler (Senckenberg), Martin Müller (Senckenberg), Dagmar Triebel (SNSB), Markus Weiss (SNSB)

Entschuldigt: Carola Söhngen (DSMZ), Lothar Menner (Senckenberg), Alexander Kroupa (MfN), Tanja Weibulat (SNSB)

Aufgrund diverser Verspätungen beginnt die Sitzung um 11.30Uhr.

TOP1 Begrüßung, Organisatorisches, Verabschiedung der Tagesordnung

A. Güntsch begrüßt alle Teilnehmer. Die Tagesordnung wird unverändert angenommen.

TOP2 Neuerungen in der technischen Infrastruktur von GBIF-International

J. Holetschek berichtet über seinen Besuch bei den Entwicklern im GBIF-Sekretariat in Kopenhagen im Juni dieses Jahres. Zu diesem TOP sind weitere Mitarbeiter und Mitarbeiterinnen des SNSB-IT-Zentrums anwesend.

Präsentation:

Siehe *2013-11-21 IT Kommission München GBIF Infrastruktur.pdf* im Anhang

Anmerkung zu „New for User: new filters/date last modified“: A. Güntsch bemerkt, dass generell ein Trend zu beobachten ist, den Datenanbietern nicht zu trauen, z.B. beim Parsen. Es wird stattdessen von GBIF selber z.B. geparkt, statt die Energie darauf zu verwenden die Provider zu besserem Parsing anzuhalten.

Livedemo der neuen Portalseite:

- Es gibt nur noch eine URL, www.gbif.org (vorher gab es noch eine zweite Datenportal-URL).
- Hinweis auf Tabs „Newsroom“ und „Uses of Data“
- Community & Country Pages, z.B. <http://www.gbif.org/country/DE/participation>
 - o Hier liefert z.B. „Data Publishing“ auch aktuelle statistische Informationen <http://www.gbif.org/country/DE/publishing>; Hinweis auf interaktive Karten und Filterbarkeit nach z.B. Basis of Records**, historische Darstellung vor 1900
- **Anmerkung Anton Güntsch zur „Basis of Records“ und fehlerhafte Darstellung von Deutschen Daten; Grund waren die ohne Basis of Record angegebenen FlorKart-Daten, wo mittlerweile nachgearbeitet wurde und bei der nächsten Indexierung sollte dies vorhanden sein. In ABCD war dies kein Pflichtfeld, so dass dieser Fehler bei einigen dt. Providern noch auftritt.
- Über die grüne Startseitenkarte wird perspektivisch die „Live-Indexierung“ anwählbar sein.
- Vorstellung der Volltextsuche: <http://www.gbif.org/publisher/search>
- Beispiel: „Listung der SNSB-Datasets“:
<http://www.gbif.org/publisher/0674aea0-a7e1-11d8-9534-b8a03c50a862>
- Beispiel: „Interaktive Suche in Occurrence“ <http://www.gbif.org/occurrence>, hier z. B. Übersicht nach „Taxonomic Characteristics“, div. Visualisierungs- und Suchmöglichkeiten mittels zwei Gruppen von Filtern; Filter beziehen sich entweder auf die Daten selbst oder auf die Metadaten;

- Blick auf den Einzelrecord und die hinterlegten und visualisierten Informationen (z.B. Zitierungsvorschlag). Beim Klick auf „Verbatim Version“ zeigt sich das, was vom Provider tatsächlich geliefert wurde – auch für die Selbstkontrolle nützlich. XML-Rohdaten werden jetzt im neuen Portal nicht mehr angezeigt.
- Anmerkung D. Triebel: „Institutional Code“ und „Sammlungscode“ sind nicht immer identisch; Hier muss man bei Suche und Datensätzen aufpassen.
- Vorstellung von „Bounding Box“ & „Polygon-Darstellung“
- Zukünftig werden „verdächtige“ Koordinaten über ein Log rückgemeldet werden.
- Für Datendownloads ist jetzt im neuen Portal eine Registrierung, also ein Nutzerkonto erforderlich. Es wird darin dann auch eine Übersicht aller bisher erfolgten Downloads gelistet, diese werden auch als Archiv bereitgestellt bzw. sind per URL weiterleitbar. Die vorherigen Suchen / Filter lassen sich jederzeit reproduzieren (bzw. URL lässt sich „bookmarken“).
- Übersicht der „Activities“ am Beispiel „Diatomeensammlung Weinzierl“; hier ist sichtbar, welche Daten wie oft runtergeladen wurden und welcher User sie runtergeladen hat.

A. Güntsch ergänzt, dass die Services auf die eigene GBIF-API aufgesetzt wurde und dass alle Services nun auch in eigene Produkte eingebunden werden können.

M. Müller fragt, ob sich bei den Wrappern etwas geändert hat. - J. Holetschek antwortet, dass sich in dem Bereich nichts geändert hat. Weiterhin werden BioCASE DiGiR, Tapir, und Dwc-A unterstützt.

J. Holetschek weist darauf hin, dass die Feedbackfunktion (Feedbackbutton) zur Mitteilung von Bugs, Anforderungen und weiteren Bemerkungen genutzt werden sollte. Es wird echtes Feedback durch das GBIF Entwicklerteam gegeben, welches sich über jede konstruktive Rückmeldung freut.

13.00 bis 13.40 Mittagsimbiss

TOP 3 http-URI-basierte Identifier für Sammlungsobjekte

A. Güntsch berichtet über die aktuellen Entwicklungen auf dem Gebiet „http-URI-basierte Identifier für Sammlungsobjekte“. Siehe Präsentation *Identifier_Guentsch.pdf* im Anhang. Besonders die schnelle Vorgehensweise und konstruktive Arbeit dieser Task-Force ist hervorzuheben und eine Beteiligung bzw. institutionelle Entscheidung für diesen Ansatz wird von A. Güntsch sehr empfohlen.

Diskussion / Fragen:

P. Grobe fragt nach, was genau erforderlich ist für die Erstellung einer URL?

G. Hagedorn: eine Subdomain überlegen, die dann verlinkt auf RDF; HTML, XML, URL.

Das Neue an dem Ansatz ist, auf eine zentrale Indexierung, wie bei DOI, zu verzichten und eine eigene Subdomain einzurichten. Unterscheidung zwischen physischem Objekt und Daten(satz) ist dabei wichtig. Siehe auch http://wiki.pro-ibiosphere.eu/wiki/DOI_or_LOD_or_DOI_and_LOD

Die Institution über nimmt die Verantwortung, die Identifier stabil zu halten und die zugehörigen Metadaten zur Verfügung zu stellen.

DOIs als zentrales System ermöglicht indes bessere statistische Auswertungsmöglichkeiten zur Nutzung von Identifiern.

Nachteil des DOI-Systems: Zentral gehaltene Metadaten passen z.T. schlecht zu biologischen Sammlungsdaten.

P. Grobe fragt nach der Suchbarkeit über Google? A. Güntsch: Den Datensatz resp. die URL findet man erst, wenn er veröffentlicht wird (zum Beispiel in einer elektronischen Publikation). Am BGBM wurde eine Sitemap implementiert, die das Finden und Harvesten von Sammlungsdaten (z.B.

durch Aggregatoren) ermöglicht. Das Aufsetzen eigener Suchportale schätzt A. Güntsch als ungünstig ein. Dies sollte eher von Aggregatoren gemacht werden.

G. Hagedorn berichtet, dass die MfN Bilder eine HTML-Seite (menschenslesbar) und RDF-Seite (maschinenlesbar) haben, aber die Entscheidung gegen ein eigenes Suchportal gefallen ist. Die Bilder werden zum Beispiel an Europeana geliefert und sind dann dort auffindbar.

Siehe auch http://wiki.pro-ibiosphere.eu/wiki/Best_practices_for_stable_URIs

Auf dieser Seite tragen SNSB, ZFMK ihre URI Strukturen nach, sobald diese in den jeweiligen Instituten abgestimmt sind.

Bei Interesse und weiten Fragen stehen Anton Güntsch und Gregor Hagedorn zur Verfügung.

TOP 4: Digitalisierungs-Wiki, Verwertung

M. Gleisberg und D. Triebel stellen den Stand der bisherigen Arbeiten kurz vor. Siehe http://wiki.gbif.de/web/Main_Page

Jede **Workstation** http://wiki.gbif.de/web/Digitalisierungsstationen_nach_Institutionen soll eine eigene Seite bekommen (nicht gruppiert, wie aktuell noch bei BGBM, DSMZ...) und es soll die Benennung von Kontaktpersonen überall ergänzt werden, ebenso Angaben, wann die Stationen etabliert wurden und wie lange die Stationen im Betrieb sind, Angaben wieviel Zeit dafür benötigt wird ein Bild einzuscannen.

- BGBM wird Einzelseiten erstellen und Bilder sowie Kontaktangaben nachliefern.
- DSMZ wird Einzelseiten erstellen und Bilder sowie Kontaktangaben nachliefern.
- MfN: Das Manual für 3DScan wird derzeit von Hiwi aufbereitet und steht bald im Wiki verfügbar. Allerdings ist der Satscan aber eher ein Fehlkauf gewesen, da qualitativ oftmals nicht ausreichend.
- Senckenberg: URL zu Digitalisierungsstationen ist noch immer rot; 2 Digitalisierungsstationen wären möglich. Stefan Dressler arbeitet bei Senckenberg im Mellon Projekt und wird von Martin Müller angesprochen.
- SNSB zwei Zuarbeiten vom IDES-Projekt stehen noch aus.
- ZFMK wird zwei Einzelseiten erstellen und Kontaktangaben ergänzen.

Zu klären ist nun, wozu das Wiki dienen kann. Eingangs war gedacht, eine Publikation daraus zu erstellen. Das kommt bei der aktuellen Fassung nicht mehr in Betracht. Der Detaillierungsgrad variiert noch immer sehr stark. Im Abschlussbericht werden die Arbeiten erwähnt werden. Andererseits wurden im gestrigen Knotenkoordinationstreffen über einen Antrag zur Digitalisierung gesprochen, so dass dieses Wiki eine gute Grundlage/Ergänzung wäre.

Es gibt zudem neuen Input aus München zu den **Workflows** unter http://wiki.gbif.de/web/Workflows_nach_Objekttypen. Der ergänzte Inhalt ist Konsens. Jede Institution füllt nun Inhalte für diese Beispiele nach. Der BGBM z.B. hinterlegt für die Arbeit in JACQ den Link vom eigenen Website, wo dies bereits beschrieben wurde.

GIS Funktionalität

http://wiki.gbif.de/web/Georeferenzierte_Sammlungsdaten_und_GIS-Funktionalit%C3%A4ten

ist seitens Senckenberg und SNSB komplett. Falls noch eine Institution Ergänzungen in diesem Bereich hat, kann dies ergänzt werden.

Alle bereiten bis zum Abschlussbericht (Projektende April 2014) die fehlenden Inhalte vor bzw. strukturieren nach Einzelseiten je Gerätetyp.

TOP 5 Neue Projekte und Anträge mit GBIF-D Bezug

- D. Triebel stellt das Projekt „Flora von Bayern“ vor, siehe <http://wiki.bayernflora.de> bzw. offizielle Pressemitteilung vom 25.11.2013 <http://www.stmug.bayern.de/aktuell/presse/detailansicht.htm?tid=27833>
- D. Triebel berichtet kurz vom Knotenkoordinationstreffen am Vortag in Braunschweig und dem Vorschlag der SNSB für eine GBIF-D 3. Phase
- SNSB (Triebel/Melzer) haben einen plakativen Vorschlag für einen gemeinsamen Folgeantrag im Bereich Digitalisierung „3 Mio (oder 3,5 Mio) neue Sammlungsobjekte aus den GBIF-D Institutionen mit Herkunft Deutschland über GBIF zugänglich – ein 5-Jahres Projekt“ vorgestellt (siehe „Skizze-GBIF-D-Digitalisierung Sammlungen Deutschland_2013-11-22.pdf“ im Anhang); d.h. jede Institution übernimmt Verantwortung für 500.000 Objekte. Der klare Bezug zu Deutschen Belegen ist gestern aber nicht Konsens gewesen. Auch müsste die Forschungsrelevanz des Ansatzes anhand konkreter Themen betont werden. Eine Kostenkalkulation dazu wurde vorgestellt und auch Minimalanforderungen der Metadaten. Bei geschätzten 2Mio EUR pro Institut wäre dies ein Gesamtvolumen von 12 Mio. (inkl. Görlitz 14 Mio.).
 - o Der Vorschlag in der Form wurde abgelehnt. Stattdessen wurde eine Gesamt-Digitalisierung diskutiert. Jetzt wird nach Sammlungsgruppen vorgegangen und ein 2-seitiges Proposal bis 16.12.2013 an die GBIF-D Koordinationsstelle gesandt, in dem eine Abschätzung zur Gesamtdigitalisierung beschrieben wird.
- M. Gleisberg ergänzt, dass es vor zwei Wochen einen Herbarkustoden-Workshop am BGBM in Berlin gab (siehe <http://www.gbif.de/botanik/Herbarkustodenworkshop>). Hieraus gibt es auch eine Ansammlung von Institutionen, die derzeit die Möglichkeit eines breit und institutionell getragenen Vor-Antrags zur Herbardigitalisierung eruiert und zeitnah realisieren will.
- G. Hagedorn und P. Grobe berichten von Gesprächen des „WGL Leibnitz“ (3 Institutionen haben dazu miteinander gesprochen: ZFMK, SENCKENBERG, MfN), der Impuls kam vom BMBF; Tendenz der drei Gesprächspartner ist, hier mit den verbleibenden GBIF-Knoteninstitutionen gemeinsam voranzugehen
 - o Der Ansatz „Alles digitalisieren“ wurde seitens BMBF so nicht angenommen; es soll ein konkreter und aktuell relevanter Forschungsansatz der Bezug sein, oder das Oberthema soll naturschutzrelevant sein (z.B. arbeitet BfN auf FFH Basis, d.h. nur diese Listen/Arten werden erfasst).
 - o Andererseits wurde soviel Geld in Aussicht gestellt, um eine Infrastruktur auf zu bauen, an welche sich weitere Projekte außerhalb der WGL anschließen könnten.
 - o Nachhaltigkeit: Der Ansatz des SNSB-Vorschlags, die Kuratoren von Anfang an und während des gesamten Projektes mit einzubeziehen ist in der Folgewirkung einerseits akzeptanzfördernd und andererseits auch nachhaltig, besonders was die Folgebetreuung der Daten betrifft.
 - o G. Hagedorn: Naturalis hat gute Erfahrungen mit zusätzlichem Personal gemacht, welches den Kuratoren direkt zugeordnet wurde.
- Es wird diskutiert, ob man IPBES als Vorgabe für Forschungsfragen hinzuziehen sollte:
 - o M. Gleisberg berichtet von der Teilnahme an dem 2. Internationalen IPBES-Forum in Bonn (06./07.11.2013, organisiert von BMU und NeFo, siehe <http://www.biodiversity.de/index.php/de/ipbes/4071-2-nationales-forum>), das als Vorbereitung auf das internationale IPBES Treffen in Antalya (09.-14.12.2013, siehe <http://www.ipbes.net/plenary/registration-ipbes-2.html>) diente. In Bonn wurde u.a. in mehreren Kleingruppen der Workplan für Antalya diskutiert und das Briefing der Deutschen Delegation durch die anwesenden Wissenschaftler vorbereitet. M. Gleisberg hat an dem Workshop Modellierung teilgenommen, dieser hatte noch den direktesten GBIF-Bezug. Der Aspekt „Modellierung“ und „Gap Analysis“ sollte in möglichen Anträgen erwähnt werden und gegebenenfalls sollte Rücksprache mit den Experten und Modellierern genommen werden. Was wird genau gebraucht, denn eine gezielte Mobilisierung ist möglich. NeFo erstellt derzeit eine solche

Expertenliste.

- Die Digitalisierung verschiedener Organismengruppen wird unterschiedlich teuer sein. (kann man durch verschiedene Digitalisierungstiefe mehr oder weniger ähnlich halten; D.T.).
- G. Hagedorn/P. Grobe: Die Forschungsfragen, die bedient werden, sollten trotzdem den Institutionen noch eine gewisse Breite in der Erfassung bieten. Eine ausgewogene Bilanz zwischen „supply- und demand driven“ ist erforderlich.
- Insgesamt ist es eine zweigeteilte Aufgabe: Inventur und Kostenschätzung
- Siehe auch Protokoll der Knotenkoordinatoren vom 20.11.2013 (noch in Vorbereitung).

TOP 6: Abstimmungen zur weiteren Arbeit der IT-Kommission

- **Weitere Arbeit der IT-Kommission**

Bis zum Ablauf der kostenneutralen Verlängerung (Projektende 30.04.2014) sind die Arbeiten im Wiki abzuschließen (siehe TOP 4).

Anton Güntsch schlägt vor, dass diese Gruppe aufrechterhalten bleibt und sich bei Bedarf treffen sollte. Das sollte aber mindestens einmal im Jahr ein physisches Treffen sein. Peter Grobe konkretisiert, dass im Herbst jeweils ein physisches Treffen und im Frühjahr eine Telekonferenz stattfinden sollte. Den Vorschlägen wird zugestimmt. Die physischen Treffen sollten nach Möglichkeit am Folgetag des Knotenkoordinatorentreffens stattfinden, da mittlerweile nahezu die gleichen Personen in beiden Gremien sitzen.

- **Sprecherwahl**

- Aktuell sind Anton Güntsch Sprecher und Gregor Hagedorn Stellvertretender Sprecher der GBIF-D IT Kommission.
- Per Akklamation wird die Wiederwahl von Anton Güntsch als Sprecher und Gregor Hagedorn als Stellvertreter bestätigt.

16:00 *Ende des Treffens*

Für das Protokoll: Maren Gleisberg, Anton Güntsch