

Protokoll

Arbeitstreffen „IBF-Informatiker“ zum DFG-Projekt „Aufbau eines Informationsnetzes für biologische Forschungsdaten von der Erhebung im Feld bis zur nachhaltigen Sicherung in einem Primärdatenrepositorium“

Zeit: 21. April 2009, 11.00–18.00 Uhr

Ort: Botanische Staatsammlung München, Arbeitszimmer
D. Triebel, Zi. 232

Teilnehmer:

Botanische Staatsammlung München: Dr. Dagmar Triebel
Dr. Markus Weiss
Dieter Neubacher

Universität Bayreuth, Ang. Informatik: Tobias Schneider
Bernhard Volz

1 Kauf der ersten Handhelds

Das erste Mobiltelefon für Dr. Alexandra Kehl aus der Arbeitsgruppe von Prof. Rambold wurde über Tobias Schneider bezogen und mit dem Prototyp der Software bespielt. Die Testversuche können am 29.04. beginnen. Das nächste Gerät (nach Rücksprache mit Jürgen Klotz: HTC Touch HD) soll in Regensburg von Jürgen Klotz aus der Arbeitsgruppe von Prof. Poschlod bestellt werden und nach Bayreuth zur Softwareinstallation gesendet werden (Lieferadresse = Bayreuth, Rechnungsadresse = Regensburg). Als weiteres Gerät soll ein Mobiltelefon in Bayreuth bei der Arbeitsgruppe von Prof. Jablonski angeschafft werden. Das SNSB wartet mit dem Kauf eines neuen Gerätes auf neue Releases.

2 Identifier

Zur Organisation des Verleihwesens in Bibliotheken (und der Zitierbarkeit von publizierter Literatur) werden global unique identifier GUIDs nach dem DOI-Standard verwendet. Eine entsprechende Standardisierung wird auch in DiversityCollection (auch DiversityTaxonNames etc.) benötigt, um Datensätze zitierbar zu machen. Dazu wird ein System benötigt, das auch eine vergebene GUID wieder auflösen (Resolve) kann und von einer zentralen Stelle verwaltet wird (Langfristige Auflösbarkeit aus sekundären Datenbanken). Da DOIs kostenpflichtig sind, bietet sich die Verwaltung über das URN-System (unified resource name) an wie er beispielsweise von der Deutschen Nationalbibliothek verwendet wird.

Da GUIDs als PK aufgrund der Größe im Vergleich zu Integer immer mit einer Performanz-Einbuße verbunden sind, soll dies zunächst in der Datenbank und insbesondere auf dem Handheld getestet werden. Auf dem Handheld wäre ein deutlicher Performance Verlust kaum tolerierbar.

Da normale GUIDs nicht aufsteigend vergeben werden, ist die Erstellung und das Arbeiten mit dem zugehörigen Index mit einem zusätzlichen Performance-Verlust verbunden. Daher sollen auch sequentielle GUIDs vergleichend untersucht werden. Ob diese auf einem PDA wegen evtl. fehlender MAC-Adresse zur Verfügung stehen muss noch eruiert werden.

3 Datenbankschema

- Datensätze in DiversityMobile und DiversityCollection werden künftig an Hand einer GUID als Primärschlüssel identifiziert. [Anmerk. M. Weiss: Dies betrifft zunächst alle Tabellen die vom Replikationsabonnenten beschrieben werden können. Inwieweit auch nicht replizierte Tabellen einbezogen werden, sollte noch diskutiert werden.]
- Zur Erfassung von Punktdaten im Gelände wurde ein Feld für Koordinaten in der Tabelle Specimen bzw. Identification Unit vorgeschlagen. [Anmerk. M. Weiss : Mit den jetzt vorhandenen Strukturen in der Datenbank ist dies bereits mittels eines eigenen CollectionEvents für ein Specimen bzw. Identification Unit möglich.]
- In DiversityMobile wird Observation Only auf Specimen-Ebene als Default-Wert eingeführt [Anmerk. M. Weiss: Observation Only wird in der Datenbank auf der Ebene des Identification Unit angegeben. Ist ein Problem?].
- Bernhard Volz erweitert die Beschreibung von möglichen Identifiern im Projekt-Wiki.
- Die Feldbezeichnungen (Nomenklatur, Begrifflichkeit) von DiversityMobile orientieren sich am Namensraum des Diversity Workbench Frameworks.

4 Use Cases zur Datenaufnahme

Um den Anwendern und den Programmierern von DiversityMobile eine Handhabe zur korrekten Speicherung der Datensätze in DiversityCollection/-Mobile zu geben, werden die einzelnen Anwendungsfälle aus Regensburg und Bayreuth im Projektwiki veröffentlicht (insbesondere sollen auch Fragen in Bezug auf Konsistenz und Konsequenzen einer Speicherungsentscheidung diskutiert werden). Die konkrete Ausarbeitung der Fragestellungen übernimmt Tobias Schneider – die Übersetzung der Anwendungsfälle in das Datenbankschema von DiversityCollection übernimmt Dr. Markus Weiss. Hierbei werden Regeln in Bezug auf die mehrfache Verwendung identischer Datensätze aufgestellt.

5 Diagramme zum Datenfluss

Das aktuelle Diagramm zum IBF-Datenfluss wird von Dr. Alexandra Kehl zusammen mit Tobias Schneider in Hinblick auf die eBiosphere-Konferenz in London überarbeitet. Im Vergleich zu der aktuellen Darstellung sollen technische Aspekte organisatorischen in Bezug auf die Darstellung vorgezogen werden.

6 Aspekte zur Synchronisation zwischen Handheld und stationären Projektdatenbanken

Ab 29.04. werden Testdatensätze von Dr. Alexandra Kehl auf ihrer Versuchsfläche aufgenommen und diese in eine Datenbank der Angewandten Informatik IV überspielt. Ist die Synchronisation zwischen Mobilgerät und Datenbank erfolgreich, wird eine Übertragung der Daten nach München vorgenommen. Da sich die Synchronisation in der Testphase befindet, werden die Daten von Dr. Alexandra Kehl parallel dazu in der traditionellen Art aufgenommen und nach München überspielt. Die Synchronisation soll dann nach Abschluss der Testphase zwischen dem Mobilgerät und der Datenbank des SNSB erfolgen. Lokale Datenbanken werden zunächst nicht unterstützt, allerdings wird eine Exportfunktion der Daten aus DiversityCollection von Dr. Markus Weiss (-> tab separated text) zu Verfügung gestellt. Die Möglichkeit einen Arbeitsnamen zu ersetzen muss sowohl in DiversityMobile wie in DiversityCollection gegeben sein und bei der nächsten Synchronisation berücksichtigt werden. Die IBF Datenbestände werden an der SNSB über Projekte verwaltet (DiversityProjects). Derzeit wurden 5 Projekte in DiversityCollection eingerichtet (IBFecocoll, IBFfungicoll, IBFlichenscoll, IBFplantscoll, IBFmergccoll].

7 Datenfluss von Bilddaten

Bilddaten müssen im Zuge der Synchronisation mit nach München übertragen werden. Dazu werden Bilder mit einer eigenen GUID versehen, welche auch als Dateiname fungiert. Referenzen in der Datenbank, die den Pfad enthalten, müssen im Zuge der Synchronisation angepasst werden. Hierbei ist eine Bildübertragung mit der Pfadanpassung eine Transaktion. Für die Organisation der Pfade in München (nach Datum, Projekt, ...) übernimmt Dr. Markus Weiss die Verantwortung. Die Entwicklungen zum Datenfluss beim BildScan-Projekt an der BSM, koordiniert von D. Neubacher, sollen berücksichtigt werden (d.h. Automatisierungen beim Datentransfer und bei den Bildkonvertierungen).

8 BioCase-Wrapper

Die DiversityCollection verfügt über einen Export in ABCD und kann somit den BioCase-Standard unterstützen.

9 DaltOn

DaltOn ermöglicht die Semantische Integration von Daten aus unterschiedlichen Datenstrukturen und die Konvertierung von Daten in verschiedene Formate. Da im

Moment in DiversityMobile mit einem zur DiversityCollection kompatiblen Datenmodell gearbeitet wird, wird Dalton zunächst nicht für die Synchronisation benötigt. Denkbar wäre ein Einsatz zur Konvertierung von Specify-Daten nach DiversityCollection (und umgekehrt) sowie von SysTax nach DiversityCollection.

10 Verschiedenes

- Zum GI Workshop-Biodiversitätsinformatik wird von der IBF-Gruppe ein Paper eingereicht, das auf dem Paper zur Konferenz in Mailand basiert.
- Als Alternative zu dem in München geplanten Wechsel von CVS nach Subversion soll u. U. auf den in Bayreuth verwendeten Jazz-Team-Server (IBM rational; IBM educational initiative) gewechselt werden, Denkbar ist dabei die Nutzung der Installation in Bayreuth oder die Einrichtung eines Jazz-Team-Servers in München (kostenfrei?). Dies soll zunächst mit Dieter Neubacher besprochen werden, der zu diesem Punkt nicht mehr anwesend war. Der zentrale Server für die Ablage des Codes könnte also, um einen späteren Umzug der Daten zu vermeiden, am SNSB IT-Zentrum aufgesetzt werden (auf BSM5, Windows 2008 Server).
- Die mündliche Kommunikation zwischen SNSB und UBT sollten wir über Skype organisieren (siehe <http://www.skype.com/intl/de/>).