

Erfahrungen mit Importen in die IMDAS-Datenbank

Dipl.-Biol. Florian Raub
Staatliches Museum für Naturkunde
Karlsruhe

Inhalt

1. Background: Naturkundliche Sammlungen
2. Sammlungen des SMNK in IMDAS
3. Anforderungen an eine Datenbank für naturkundliche Datenbestände
4. Datenimporte in Datenbanksysteme
5. Exkurs: Biologischer Systematik-Thesaurus
6. Import in IMDAS
7. Verwendung und Verfügbarmachung von Daten in IMDAS
8. Fazit

Naturkundliche Sammlungen...

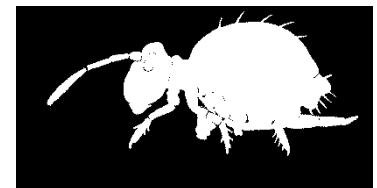
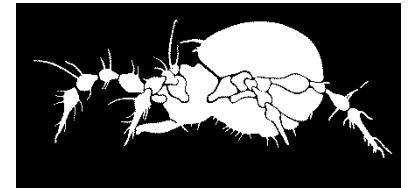
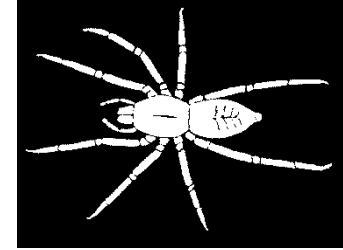
- ... sind zum Teil sehr alt und die Daten sind daher in den unterschiedlichsten Datenformaten und in unterschiedlichster Qualität vorhanden.
- ... sind heterogen was den (monetären) Wert angeht.
 - Tax. Sammlungen meist ohne kommerziellen Wert aber arbeitsintensiv
 - Paläontologische Sammlungen oft mit sehr hohem kommerziellen Wert
 - Einzigartige Objekte mit hohem (wissenschaftl.) Wert (ausgestorbene Tiere)
- ... bestehen aus sehr vielen Einzelobjekten mit teilweise umfangreichen und sehr heterogenen Daten und Metadaten.
 - Ortsdaten
 - Ökologische Metadaten
 - Sammlerdaten
 - Projektdaten
- ... sind „Forschungssammlungen“ mit denen multinational und in den verschiedensten Bereichen gearbeitet wird.
 - Taxonomische Belegsammlungen → Typusmaterial in tax. Sammlungen
 - Ökologische Sammlungen aus Forschungsprojekten
 - Statistische Auswertungen, Metaanalysen großer Datenbestände
 - Ausstellungsarbeit mit Sammlungsobjekten → „Material in Bewegung“

Naturkundliche Sammlungen...



Sammlungen des SMNK in IMDAS

- ~ 20.000 Säugetiere (Mammalia)
- ~ 4.600 Vögel (Aves)
- ~ 4.600 Spinnen (aus Deutschland)
- ~ 6.000 Hornmilben (Oribatida)
- ~ 4.500 Springschwänze (Collembola)
- ~ 23.000 Pilze (in Vorbereitung)
- ~ 12.000 Gottesanbeterinnen (Mantodea)
- Mineralien & paläontologische Datensätze in Vorbereitung



Anforderungen an eine Datenbank für naturkundliche Datenbestände

- Sammlungsverwaltung
 - Inventarisierung, Ausleihvorgänge & Bewegung der Objekte innerhalb des Hauses
- Flexible Datenfeldstrukturen zur Anpassung an verschiedene Metadaten
- Flexibler Datenimport
 - Excel-Listen
 - Relationale Datensätze
- Einfache Möglichkeit der Georeferenzierung (z.B. über GoogleMaps?)
- Anbindung an öffentliche Thesauri (Catalogue of Life, Index Fungorum,...)
- Problemlose Übergabemöglichkeiten an Biodiversitätsdatenportale (via Datenwrapper oder direkter Übergabe)
 - GBIF
- Integrierte Auswertungsmöglichkeiten
 - Verbreitungskarten
 - Statistische Analysen
- Zurverfügungstellung der Datensätze (in Informationssystemen/Portalen)
 - Für die Öffentlichkeit (Museumshomepage)
 - Für Wissenschaftler/Institutionen (intern oder extern z.B. zur Auswertung im Rahmen von Metaanalysen) → Metadaten extrem wichtig!
- Export von (Teil-)Datensätzen (und Metadaten) zur Analyse in externen Programmen

Datenimport in Datenbanksysteme

- Datenimport in ein neues (relationales) Datenbanksystem als arbeitsaufwendiger Schritt
 - Inhaltliche Arbeit:
 - Systematik-Thesaurus erzeugen
 - Bei alten Sammlungen neue Systematik recherchieren
 - Auswahl eines geeigneten taxonomischen Thesaurus
 - » je nach Gruppe fachlich sehr problematisch!

Exkurs: Biologischer Systematik- Thesaurus

- Systematik Hauptkategorien (nach Cavalier-Smith, 1998):

Domäne → Reich → Stamm → Klasse → Ordnung → Familie → Gattung → Art

– z.B. der Wolf:

Domäne: Eukaryota (Lebewesen mit Zellkern)

- Reich: Tiere – Animalia

– Stamm: Chordatiere – Chordata

» Klasse: Säugetiere – Mammalia

- Ordnung: Raubtiere – Carnivora

- Familie: Hunde – Canidae

- Gattung: Canis

- Art: lupus

- Autor: Linnaeus 1758



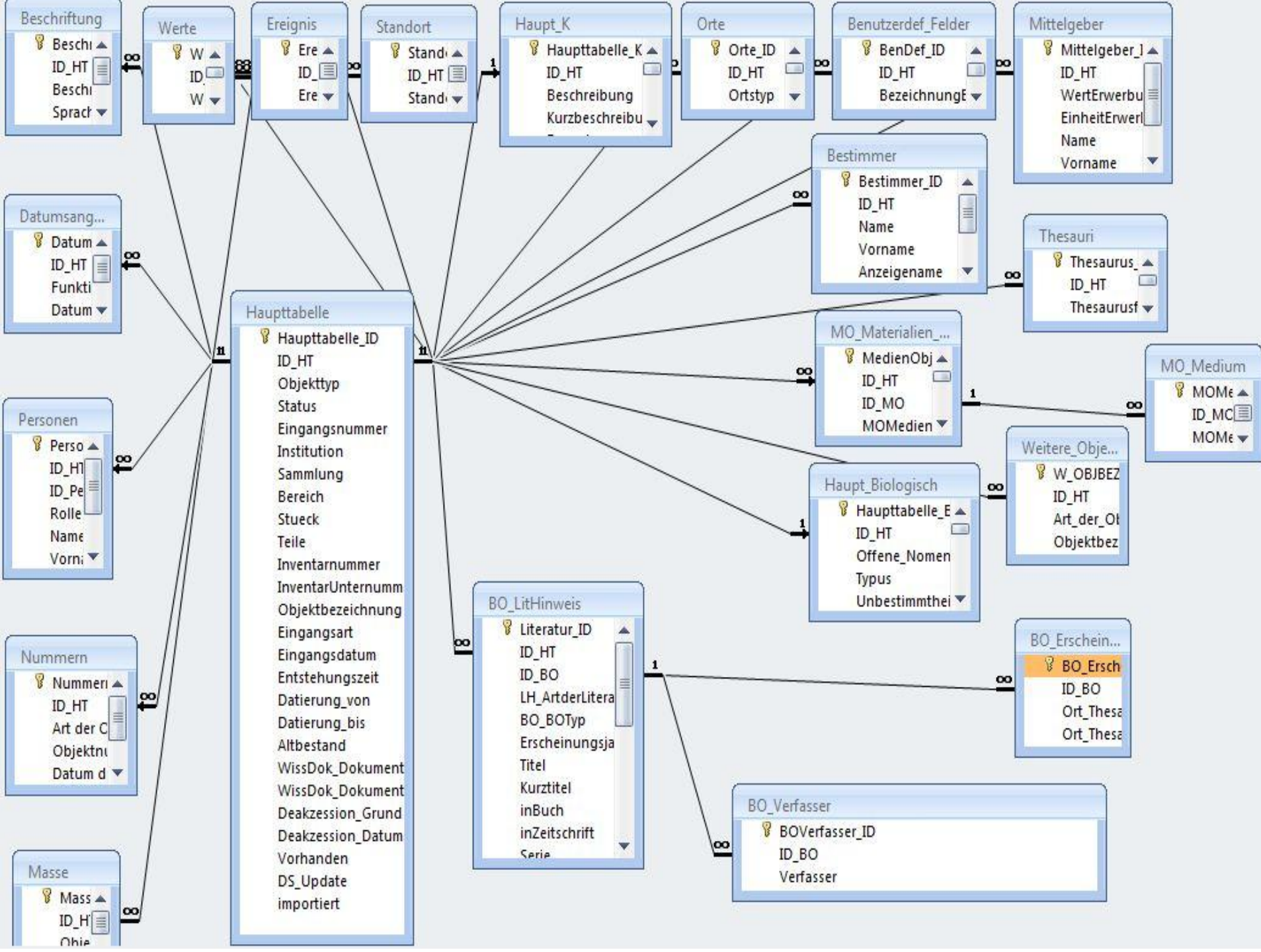
Canis lupus Linnaeus, 1758

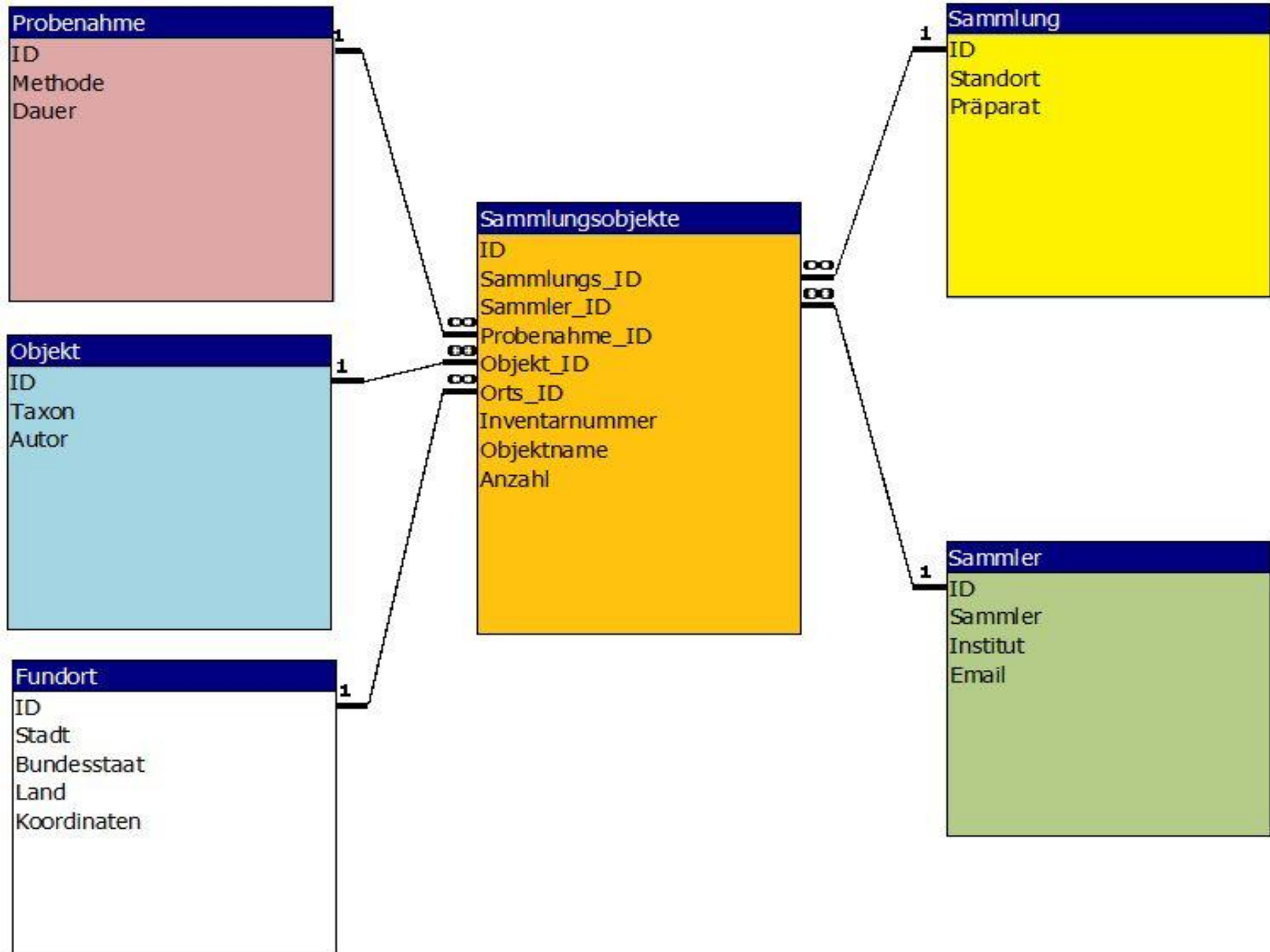
Datenimport in Datenbanksysteme

- Datenimport in ein neues (relationales) Datenbanksystem als arbeitsaufwendiger Schritt
 - Inhaltliche Arbeit:
 - Systematik-Thesaurus erzeugen
 - Bei alten Sammlungen neue Systematik recherchieren
 - Auswahl eines geeigneten taxonomischen Tesaurus
 - » je nach Gruppe fachlich sehr problematisch!
 - Daten bereinigen/berichtigen
 - Nachträgliche Recherchen (Georeferenzierung)
 - Technische Arbeit
 - Daten selektionieren (nach Qualitätsmerkmalen)
 - Daten eingeben
 - Daten für den Import vorbereiten

Import in IMDAS

- Importdatenbank aufgrund von (nicht korrekt relationalen) Dopplungen aufwendig in der Verwendung und fehleranfällig
- Excel-Import (z.B. nach einem Export aus relationalen Systemen) nicht möglich
- Thesauruserstellung und Pflege sehr aufwendig und problematisch
 - Beginnt erst bei Klasse (Pilze als Problemfall! → Fungi, Protozoa, Plantae)
 - Keine Zwischenebenen wie z.B. Abteilung verwendbar
 - Autoren (zwingender Bestandteil einer tax. Einheit!) nur als Verweis → schlechte Exportmöglichkeit
- Erstellung von restriktiven Stammdatenlisten für umfangreiche Datensätze und große Institutionen sehr zeitaufwendig (Personenstammdaten), z.T. fehleranfällig und mit wenig Datenverbesserung verbunden
 - Personenstammdaten nur Institutions-, nicht aber sammlungsspezifisch
→ häufige Doppeleingaben oder aber Fehler bei Zuordnungen unvermeidbar
- Ortsthesauruserstellung zeitaufwendig und (bei alten Datenbeständen oder außerdeutschen Datensätzen) schlecht oder gar nicht durchführbar
 - Vorgegebene Ortslisten außerhalb Deutschlands unzureichend
 - Import von Daten mit und ohne Thesaurusbezug wäre sinnvoll





Verwendung & Verfügbarmachung von Daten in IMDAS

- Wenig (statistischen) Auswertungsmöglichkeiten in IMDAS
- Kein Systematikthesaurus für Wirtsorganismen verwendbar (z.B. wichtig bei Pilzen)
- Vollständige Webschnittstelle (zur Nutzung in Homepages) als Zusatzmodul
- Keine GBIF-Anbindung vorhanden → Datenexport nötig
- Datenexport
 - (je nach Datensatzgröße) sehr zeitaufwendig
 - Nicht komplett (txt: Verweise: Autoren fehlen, xml: Systematikpfad kann nicht komplett exportiert werden)

Fazit

- IMDAS zur Inventarisierung und Verwaltung naturwissenschaftlicher Sammlungen gut geeignet
- Wissenschaftliches arbeiten mit IMDAS-Datensätzen oft umständlich bzw. (durch unerlässliche Exportvorgänge) zeitaufwendig
- Import großer (und speziell alter) Datensätze in IMDAS sehr zeitintensiv ohne großen Zugewinn an Datenqualität
- Anbindung von IMDAS an bestehende Datenbanklösungen (z.B. GBIF) bzw. Internetseiten und Informationsportale aufwendig in der Realisierung bzw. nur indirekt (über mehrere Exporte) möglich

Vielen Dank für Ihre Aufmerksamkeit!

