Better Together: Merging our knowledge about people, places, collections, and taxonomies with Wikidata

David Fichtmueller, Anton Güntsch*, Deborah L. Paul, Thierry Bourgoin, Donat Agosti, Eva Häffner

*Freie Universität Berlin
Botanic Garden and Botanical Museum Berlin

(My) brief history of collection data management

Around the year 1998:

- Data management was largely carried out locally.
- Dozens of individual software solutions.
- Restrictive publication policies.
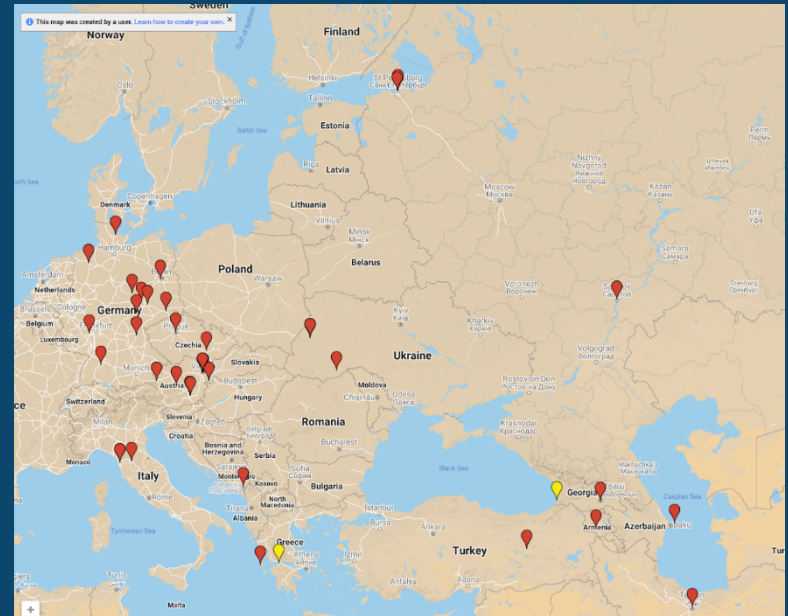- It was not generally accepted that data sharing is a good idea.

10 years later …

Software development cooperation

- Collections started to re-use existing software.
- Some collections got together to develop software themselves.
- Completely new systems but also adaptations of existing open source solutions.

Today …

Shared databases and platforms

- Example: The JACQ system.
- 50 herbarium collections using a single (!) shared database.
- Master database hosted at the Natural History Museum Vienna.
- Several synchronised copies in European Collections.

Many advantages

- Core data types are maintained jointly (e.g. scientific names, controlled vocabularies).
- Harmonised data input procedures.
- Re-use already existing data (for example in the case of duplicates).
- Shared and harmonised networking components (GBIF, BioCASe, etc.).

But …
data are still not linkable and re-usable on a global level.

Can we go further?

- Data types could be jointly maintained in the public domain and linked locally.
- Example: persons (collectors).
- Maintain biography jointly for example in Wikidata.
- Simply link the local data with the Wikidata ID.

Authority Management of People Names Workshop @ BiodiversityNext Conference 2019

Freie Universität Berlin

BG BM | Botanischer Garten & Botanisches Museum Berlin

Example: Richard Spruce (Wikidata: q1349394)

Linking Richard Spruce (q1349394) in JACQ

Stable identifier	http://herbarium.bgbm.org/object/B200097721
Collection Herb.#	B 20 0097721
Stored under taxonname	Abrodictyum cellulosum (Klotzsch) Ebihara & Dubuisson
Family	Hymenophyllaceae
Det./rev./conf./assigned	B. Zimmer (B) 2017-10-30
Ident. history	orig: Trichomanes trichoideum; G. Mettenius: P. Windisch 1994-06: Trichomanes filiforme; B. Zimmer (B) 2017-10-30:= Abrodictyum cellulosum
Isotypus for	Trichomanes filiforme J. W. Sturm
	Typified by: P.G. Windisch 1994-06
Collector	Spruce,R. 2278 (link to CETAF Botany Pilot)

Linked information for Richard Spruce (q1349394)

- Biography
- Publications
- Specimens from different collections

Advantages of the approach

- Duplicate work in data maintenance is avoided.
- Improved data quality.
- Easy integration of additional information sources.
- Creation of a common information space across collections.

Extending the approach to other data types

- Places
- Scientific names
- Taxonomic concepts
- Controlled vocabularies (e.g. habitat types)
- Collections

Much of this is already being addressed in various working groups and initiatives.

Can we perhaps go even further?

- Give up local collection management and move it into the public domain.
- Only data on storage, loans, etc. would be maintained locally.

?

Freie Universität Berlin

BG BM | Botanischer Garten & Botanisches Museum | Berlin

What do people think?

In 2019 (~30) members of a CETAF ISTC / Digitisation Working Group meeting were asked whether they found the idea of **maintaining collection data in the public domain interesting** and whether it should be investigated further.

Surprise ☺

Kontakt   a.guentsch@bgbm.org
Twitter   @aguentsch

BGBM/BDI   https://www.bgbm.org/de/wissenschaft-biodiversitaetsinformatik