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.
Example: Richard Spruce (Wikidata: q1349394)
Linking Richard Spruce (q1349394) in JACQ

<table>
<thead>
<tr>
<th>Stable identifier</th>
<th><a href="http://herbarium.bgbm.org/object/B200097721">http://herbarium.bgbm.org/object/B200097721</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection Herb.#</td>
<td>B 20 0097721</td>
</tr>
<tr>
<td>Stored under taxonname</td>
<td>Abrodictyum cellulosum (Klotzsch) Ebihara &amp; Dubuisson</td>
</tr>
<tr>
<td>Family</td>
<td>Hymenophyllaceae</td>
</tr>
<tr>
<td>Det/rev/conf./assigned</td>
<td>B. Zimmer (B) 2017-10-30</td>
</tr>
<tr>
<td>Ident. history</td>
<td>orig: Trichomanes trichoideum; G. Mettenius: P. Windisch 1994-06: Trichomanes filiforme; B. Zimmer (B) 2017-10-30: = Abrodictyum cellulosum</td>
</tr>
<tr>
<td>Isotypus for</td>
<td>Trichomanes filiforme J. W. Sturm</td>
</tr>
<tr>
<td>Typified by</td>
<td>P.G. Windisch 1994-06</td>
</tr>
<tr>
<td>Collector</td>
<td>Spruce, R. 2278 (link to CETAF Botany Pilot)</td>
</tr>
</tbody>
</table>
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.
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.
This will make history: 95% of #CETAF #ISTC 2019 participants find the idea of stopping local specimen data management and moving "everything" to @wikidata interesting. What about a pilot project with a sample set of specimens from CETAF collections? @eurotaxonomy
Kontakt  a.guentsch@bgbm.org
Twitter  @aguentsch

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