Using Wikibase as a Platform to Develop a Semantic Biodiversity Standard

David Fichtmueller
Botanic Garden and Botanical Museum Berlin
The Person

- Computer Scientist at the Botanic Garden and Botanical Museum Berlin (BGBM)
- Wikipedia Contributor since 2004
- Wikidata Contributor since 2014
- Co-Administrator of a MediaWiki farm with >80 wiki instances
The Person

- Working on Standards Development
- New Version of the ABCD Standard
The ABCD Standard

- Access to Biological Collection Data
- XML Schema
- >1800 Concepts (XPaths)
The Challenge

• Convert it to a semantic standard (OWL Ontology) with the same expressiveness
• Keep existing documentation, restrictions, connections
• Collaborative Platform for many users
The Solution

WikidataCon 2019, 25-26 October, Berlin
The Setup

• Installation without Docker Image

**BIG MISTAKE! Don’t do this!**

(we only did it because the Docker Image didn’t exist back then)

• Installation of Docker for Query Service
• Installation of QuickStatements (manually)
Warning: Possible Confusion

- The ABCD Standard will be modelled as Classes and Properties (Object Properties and Datatype Properties)
- Both are expressed as Wikibase Items
- ABCD Properties are described using Wikibase Properties
# The Properties

- **25 Wikibase Properties to describe the Items**

<table>
<thead>
<tr>
<th>Label</th>
<th>Description</th>
<th>Aliases</th>
</tr>
</thead>
<tbody>
<tr>
<td>is part of</td>
<td>in Vocabulary</td>
<td>Usage Notes</td>
</tr>
<tr>
<td>Concept Name</td>
<td>additional Description</td>
<td>Created</td>
</tr>
<tr>
<td>has range</td>
<td>in Concept Group</td>
<td>Modified</td>
</tr>
<tr>
<td>has datatype</td>
<td>Example</td>
<td>associated Property</td>
</tr>
<tr>
<td>allowed value</td>
<td>Example (localized)</td>
<td>associated Concept</td>
</tr>
</tbody>
</table>
The Import

• Extracted all XML Concepts using a Schema Parser
  Name, XPath, Documentation, ...
• Import using Quick Statements
The Editorial Process

- Manual editing to reduce the number of concepts
  - 1800+ were reduced to 543 Items
  - 64 Classes, 88 Object P., 301 Datatype P.
- Large scale edits with QuickStatements
The Validation

• Maintenance Queries for the SPARQL Interface to check consistency of modelling
• Manual Review of all concepts
The Export

• Export to OWL using a Python Script that uses the Query Service to fill templates
• Export to JSON for documentation website
• Export to a second Wikibase instance to use the concepts to describe biological data
The Result

• Success!
• Huge interest from the Biodiversity Information Standards (TDWG) Community to reuse approach for other standards
The Future

- Export restrictions as Shape Expressions (ShEx)
- Build workflows for versioning
- Switch from chimera installation to pure Docker setup
The Wish List

• Make Wikidata tools more instance-agnostic in particular: QuickStatements

• Import wikibase related templates into docker image and enable required Extension in particular: \{SPARQL2\}, \{Q\},\{P\}
The Wish List

• A tool for manual mass editing

define a list of items and properties (e.g. as SPARQL query) and allow Excel like tabular editing, quickly jumping from one item/property to another

<table>
<thead>
<tr>
<th>Item</th>
<th>Label (en)</th>
<th>Descr. (en)</th>
<th>Connected to (P123)</th>
<th>Count (P572)</th>
<th>Ext ID (P458)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item A</td>
<td>Item A</td>
<td>Lorem ipsum</td>
<td>Item X (Q484)</td>
<td>572</td>
<td>XRA</td>
</tr>
<tr>
<td>Item B</td>
<td>Item B</td>
<td>dolor sit amet</td>
<td>Item Y (Q4156)</td>
<td>745</td>
<td>MCZ</td>
</tr>
<tr>
<td>Item C</td>
<td>Item C</td>
<td>consetetur sadipscing</td>
<td>Item Z (Q841)</td>
<td>882</td>
<td>I</td>
</tr>
</tbody>
</table>
• … of this talk
• … but the beginning of a new approach to community based standards development