

## **FAIR Digital Object design for data from multiple sources (Topic 6)**

### *Aim/problem/goal*

To design a workflow to create semantically enhanced FAIR Digital Objects that can interconnect disparate biodiversity data in the different research infrastructures within a coherent structure in the near future. Moreover, to provide a JSON-LD representation of an Open Digital Specimen (ODS) and an automated way to validate the structure of data against the ODS standard.

### *Method*

An automated workflow was developed to create semantically enhanced Digital Objects validated against the new modelling framework created for ODS modelling. To this end, a [Wikibase environment](#) was set up and customised to develop the data types and properties for ODS. The Wikibases enabled the transformation of data types and properties from the DiSSCo's Modelling Framework into Shape Expressions (ShEx). A Digital Specimen is expected to satisfy this ShEx schema. Based on the ShEx schema, a workflow is designed to validate biodiversity data against the data model.

### *Results*

- A framework and recipe to label schema and align terms.
- A new FAIR Digital Object schema and corresponding data types.
- A model/ontology based on a simple example of a Digital Specimen.
- A fully automated workflow from model to ShEx schema in Cordra and semantic validation of new digital specimen objects.

### *Conclusion*

Even though JSON schema provides a structure, it cannot explicitly capture various aspects of the data such as type definition, constraints. RDF statements also can be incomplete or missing. It is important to explicitly articulate the schema (Kellou-Menouer et al. 2021). ShEx in this regard is more powerful than simple schema validation and can also help with the presentation of the semantics for both humans and machines. The use of Wikibase as a modelling framework has advantages for creating a future ODS standard compliant with Biodiversity Information Standards (TDWG) requirements due to better tracking of versioning.

### *References*

- Kellou-Menouer, Kenza, Nikolaos Kardoulakis, Georgia Troullinou, Zoubida Kedad, Dimitris Plexousakis, and Haridimos Kondylakis. 2021. 'A Survey on Semantic Schema Discovery'. *The VLDB Journal*, November. <https://doi.org/10.1007/s00778-021-00717-x>.