

Biodiversity Literature Repository: Building the customized FAIR repository by using custom metadata

Alexandros Ioannidis-Pantopikos[‡], Donat Agosti[§]

[‡] Zenodo / CERN, Meyrin, Switzerland

[§] Plazi, Bern, Switzerland

Corresponding author: Alexandros Ioannidis-Pantopikos (a.ioannidis@cern.ch)

Abstract

In the landscape of general-purpose repositories, [Zenodo](#) was built at the European Laboratory for Particle Physics' ([CERN](#)) data center to facilitate the sharing and preservation of the long tail of research across all disciplines and scientific domains. Given Zenodo's long tradition of making research artifacts [FAIR](#) (Findable, Accessible, Interoperable, and Reusable), there are still challenges in applying these principles effectively when serving the needs of specific research domains.

[Plazi](#)'s biodiversity taxonomic literature processing pipeline liberates data from publications, making it FAIR via extensive metadata, the minting of a [DataCite](#) Digital Object Identifier (DOI), a licence and both human- and machine-readable output provided by Zenodo, and accessible via the [Biodiversity Literature Repository](#) community at Zenodo. The deposits (e.g., [taxonomic treatments](#), [figures](#)) are an example of how local networks of information can be formally linked to explicit resources in a broader context of other platforms like [GBIF](#) (Global Biodiversity Information Facility).

In the context of biodiversity taxonomic literature data workflows, a general-purpose repository's traditional submission approach is not enough to preserve rich metadata and to capture highly interlinked objects, such as taxonomic treatments and digital specimens. As a prerequisite to serve these use cases and ensure that the artifacts remain FAIR, Zenodo introduced the concept of [custom metadata](#), which allows enhancing submissions such as figures or taxonomic treatments (see as an example the treatment of [Eurygyrus peloponnesius](#)) with custom keywords, based on terms from common biodiversity vocabularies like [Darwin Core](#) and [Audubon Core](#) and with an explicit link to the respective vocabulary term.

The aforementioned pipelines and features are designed to be served first and foremost using public Representational State Transfer Application Programming Interfaces (REST APIs) and open web technologies like [webhooks](#). This approach allows researchers and

platforms to integrate existing and new automated workflows into Zenodo and thus empowers research communities to create self-sustained cross-platform ecosystems. The [BiCIKL project](#) (Biodiversity Community Integrated Knowledge Library) exemplifies how repositories and tools can become building blocks for broader adoption of the FAIR principles.

Starting with the above literature processing pipeline, the concepts of and resulting FAIR data, with a focus on the custom metadata used to enhance the deposits, will be explained.

Keywords

repositories, preservation, findability

Presenting author

Alexandros Ioannidis-Pantopikos

Presented at

TDWG 2021

Funding program

The BiCIKL project receives funding from the European Union's Horizon 2020 Research and Innovation Action under grant agreement No 101007492

Conflicts of interest