

Making Schemas and Mappings Available and FAIR: A metadata and schema crosswalk registry from the FAIRCORE4EOSC project

Tommi Suominen[‡], Joonas Kesäniemi[‡], Hanna Koivula[‡]

[‡] CSC - IT Center for Science, Espoo, Finland

Corresponding author: Hanna Koivula (hanna.koivula@csc.fi)

Abstract

Community standards like the Darwin Core (Darwin Core Task Group 2009) together with semantic artefacts (controlled vocabularies, ontologies, thesauri, and other knowledge organisation systems) are key building blocks for the implementation of the FAIR (Findable, Accessible, Interoperable, Reusable) principles (Wilkinson et al. 2016), specifically as emphasized in the Interoperability [principle I2 “\(Meta\)data use vocabularies that follow FAIR principles”](#). However, most of these artefacts are actually not FAIR themselves (Le Franc et al. 2020).

To address this, the [FAIRCORE4EOSC project \(2022-25\)](#) is developing a [Metadata Schema and Crosswalk Registry](#) (MSCR) that will allow registered users and communities to create, register and version schemas and crosswalks that all have persistent identifiers (PIDs). The published content can then be searched, browsed and downloaded without restrictions. The MSCR will also provide an API to facilitate the transformation of data from one schema to another via registered crosswalks. It will provide projects and individual researchers with the possibility to manage their metadata schemas and/or relevant metadata schema crosswalks. The schema and crosswalks will be shared with the community for reuse and extension supported by a proper versioning mechanism.

The registry tool will facilitate better interoperability between resource catalogues and information systems using different (metadata) schemas and encourage organisations and especially researchers to share their metadata interoperability by publishing the metadata crosswalks used in their workflows, which are currently not visible (FAIRification). By providing an easy-to-use graphical user interface (GUI) for creating crosswalks, the GUI will attract users currently relying on project-specific solutions.

Keywords

semantic artefacts, linked open data, interoperability, persistent identifiers, European Open Science Cloud

Presenting author

Hanna Koivula

Presented at

TDWG 2023

Hosting institution

CSC - IT Center for Science

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Darwin Core Task Group (2009) Darwin Core. Biodiversity Information Standards (TDWG). <http://www.tdwg.org/standards/450>. Accessed on: 2023-6-30.
- Le Franc Y, Parland-von Essen J, Bonino L, Lehv slaiho H, Coen G, Staiger C (2020) D2.2 FAIR Semantics: First recommendation. <https://doi.org/10.5281/zenodo.5361930>
- Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1). <https://doi.org/10.1038/sdata.2016.18>