International FAIR Digital Objects Implementation Summit 2024

Extended Abstracts

https://doi.org/10.52825/ocp.v5i.1179

Published: 18 Mar. 2025

Implementing FAIR Semantic Mappings Leveraging on RO-Crate

Jonas Grieb[1], Claus Weiland[1], Alexander Wolodkin[1], Julian Lopez Gordillo[2], and Sharif Islam[2]

[1]Senckenberg – Leibniz Institution for Biodiversity and Earth System Research, Germany

[2]Naturalis Biodiversity Center, Netherlands

*Correspondence: Jonas Grieb, jonas.grieb@senckenberg.de

Abstract. The mapping.bio platform for the curation of FAIR semantic mappings is presented and the implementation of the storage of the mappings as “webby” FAIR Digital Objects based on established standards is described.

Keywords: FAIR Digital Object, Biodiversity Digital Twin, SSSOM, FAIR Signposting, Bio-schemas, EOSC

Implementing FAIR Semantic Mappings Leveraging on RO-Crate

Semantic mappings, i.e. sets of correspondences or matches between terms from semantic artifacts (controlled vocabularies, thesauri, ontologies), are essential for cross-domain interoperability and reuse of data [1]. A recent design study for a "Flexible Semantic Mapping Framework” [2] proposes both a data model, rooted in the FAIR Digital Object (FDO) approach [3] and a corresponding service infrastructure to foster sharing and publishing of mappings employing open registries.

Building on SEMAF’s design, a web-based service for the creation, curation, documentation and publication of FAIR semantic mappings is under development within the framework of Horizon Europe Project “Biodiversity Digital Twin for Advanced Modelling, Simulation and Prediction Capabilities" (BioDT - https://biodt.eu): Mapping.bio (https://mapping.bio) realizes a web service and repository based on the digital object middleware CORDRA (https://www.cordra.org) to read semantic artifacts, visualize them, add mappings as graphical connections and store the mappings as detached RO-Crates [4]. As opposed to attached RO-Crates (which are present in a file-system context) in a detached RO-Crate, all data need to be defined as Web-based data entities with absolute references. Using RO-Crates in mapping.bio enables a clear implementation path for “webby FDOs” leveraging on (i) a model for the web topology of resources ("links") and their type ("link relation types", RFC 8288 - https://datatracker.ietf.org/doc/html/rfc8288) following the FAIR Signposting Profile (https://signposting.org/FAIR), (ii) structured metadata to define the model layer based on schema.org (https://schema.org), and (iii) a comprehensive description of the actual semantic mappings compliant with the Simple Standard for Sharing Ontological Mappings [5]. This is illustrated in Fig. 1: Individual mappings are stored as a MappingSet according to the SSSOM standard and encapsulated in an RO-Crate document. A persistent identifier (PID) is minted for the record in the form of a handle. Kernel metadata elements which are important for cross-domain resource discoverability (e.g., the type, license, name and information on contextual entities like the creator(s) of the resource) are stored both in the RO-Crate (using vocabulary

from schema.org) and as key-value pairs in the PID record. Thus, machine agents are enabled to retrieve the metadata and SSSOM-conformant mappings either by analyzing the elements in the PID record or by resolving the PID, which returns typed links according to the FAIR Signposting Profile pointing to the machine-readable RO-Crate serialization of the resource. This implementation is a proposal for “webby” FDOs. Future work will include the detailed assessment of compliance with the FDO specification [6].

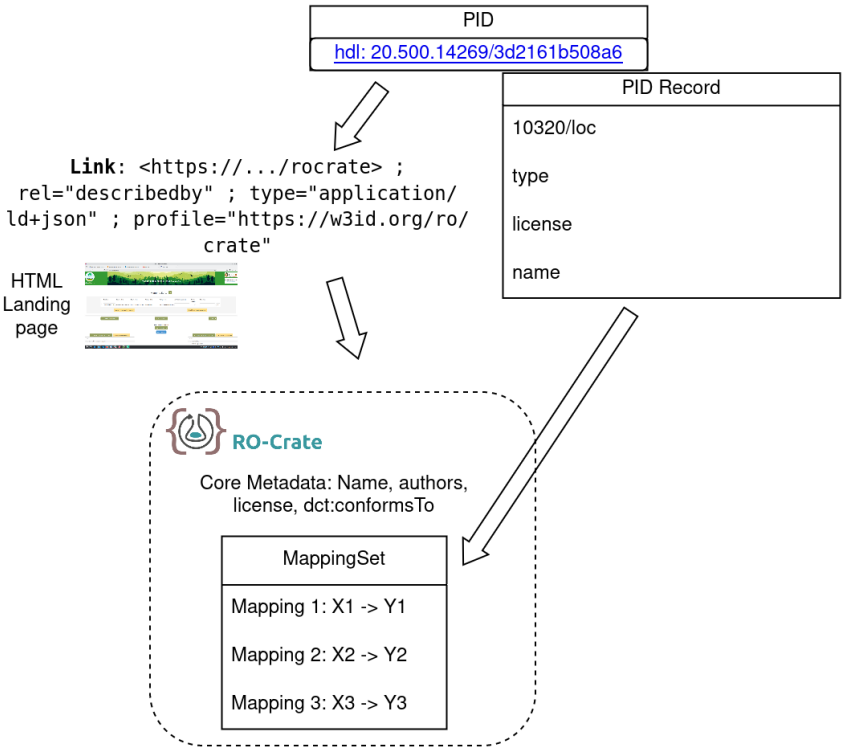

***Figure 1**: Sketch of how the standardized metadata is provided to machine agents either by resolving the PID and following typed links (FAIR Signposting) to the mappings embedded in an RO-Crate or by reading the kernel metadata in the PID record.*

With regard to the functional context of Biodiversity Digital Twins, an additional focus is placed on opening up so-called pragmatic mappings, which means mappings that are driven by specific interoperability goals such as translations between specific observation measurements (e.g. sensor configurations) and metadata describing field observations [7].

We will present recent upgrades of mapping.bio, including an extended range of supported serialization formats, improved usability and - with regard to requirements relating to the inclusion into BioDT’s FDO layer [8] - adaptations in the data model, which encapsulates the actual mapping data conforming the SSSOM specification in an FDO ‘shell’.

Finally, we will give an outlook on the integration of mapping.bio into federative infrastructures such as the service ecosystem provided as core components for the European Open Science Cloud (https://eosc-portal.eu), notably the Metadata Schema and Crosswalk Registry (https://faircore4eosc.eu/eosc-core-components/metadata-schema-and-crosswalk-registry-mscr) and the EOSC Data Type Registry (https://faircore4eosc.eu/eosc-core-components/eosc-data-type-registry-dtr).

Data availability statement

Not applicable because the abstract describes the configuration of a technical infrastructure.

Author contributions

JG, CW: Writing - Original draft preparation, Conceptualization, Methodology, Software. AW: Software, Visualization, Writing – Review and Editing. JLG, SI: Writing – Review and Editing. CW, SI: Funding acquisition.

Competing interests

The authors declare that they have no competing interests.

Funding

This work has been supported by B*iodiversity Digital Twin for Advanced Modelling, Simulation and Prediction Capabilities (BioDT)* (Horizon Europe, GA no. 101057437). Work corresponding to CW and AW has been additionally supported by *BITS - BluePrints for the Integration of Terminology Services in Earth System Sciences* (Deutsche Forschungsgemeinschaft/DFG - GA no. 508107981). Work corresponding to JG has been additionally supported by NFDI4Earth (Deutsche Forschungsgemeinschaft/DFG - GA no. 460036893).

References

[1] Jonquet C, Dutta B, Bonino da Silva Santos L, Pergl R, Le Franc Y. (2023) Common Minimum Metadata for FAIR Semantic Artefacts. 2nd Workshop on Ontologies for FAIR and FAIR Ontologies (Onto4FAIR 2023) Jul 2023, Sherbrooke, Canada. https://hal.science/hal-04106533/document

[2] Broeder D, Budroni P, Degl'Innocenti E, Le Franc Y, Hugo W, Jeffery K, Weiland C, Wittenburg P, Zwolf CM (2021) SEMAF: A Proposal for a Flexible Semantic Mapping Framework. Zenodo https://doi.org/10.5281/zenodo.4651420

[3] Wittenburg P, Schwardmann U, Blanchi C, Weiland C (2023) FDOs to Enable Cross-Silo Work. Vol. 1 (2023): 1st Conference on Research Data Infrastructure (CoRDI) - Connecting Communities. https://doi.org/10.52825/cordi.v1i.263

[4] Soiland-Reyes S, Sefton P, Castro LJ, Coppens F, Garijo D, Leo S, Portier M, Groth P (2022) Creating lightweight FAIR Digital Objects with RO-Crate. Research Ideas and Outcomes 8: e93937. https://doi.org/10.3897/rio.8.e93937

[5] Matentzoglu N, Balhoff JP, Bello SM, Bizon C, Brush M, Callahan TJ, Chute CG, Duncan WD, Evelo CT, Gabriel D, Graybeal J, Gray A, Gyori BM, Haendel M, Harmse H, Harris NL, Harrow I, Hegde HB, Hoyt AL, Hoyt CT, Jiao D, Jiménez-Ruiz E, Jupp S, Kim H, Koehler S, Liener T, Long Q, Malone J, McLaughlin JA, McMurry JA, Moxon S, Munoz-Torres MC, Osumi-Sutherland D, Overton JA, Peters B, Putman T, Queralt-Rosinach N, Shefchek K, Solbrig H, Thessen A, Tudorache T, Vasilevsky N, Wagner AH, Mungall CJ (2022) A Simple Standard for Sharing Ontological Mappings (SSSOM). Database 2022 https://doi.org/10.1093/database/baac035

[6] Ivonne, A., Christophe, B., Daan, B., Maggie, H., Sharif, I., Thomas, J., Larry, L., Karsten, P.-. von G., Robert, Q., Alexander, S., Ulrich, S., Stian, S.-R., Strawn, G., Dieter, . van U., Claus, W., Peter, W., & Carlo, Z. (2023). FDO Forum FDO Requirement Specifications. https://doi.org/10.5281/zenodo.7782262

[7] Wolodkin A, Weiland C, Grieb J (2023) Mapping.bio: Piloting FAIR semantic mappings for biodiversity digital twins. Biodiversity Information Science and Standards 7: e111979. https://doi.org/10.3897/biss.7.111979

[8] Islam S, Koivula H, Endresen D, Kusch E, Schigel D, Addink W (2023) Data Standards and Interoperability Challenges for Biodiversity Digital Twin: A novel and transformative approach to biodiversity research and application. Biodiversity Information Science and Standards 7: e112373. https://doi.org/10.3897/biss.7.112373