International FAIR Digital Objects Implementation Summit 2024

Extended Abstracts

https://doi.org/10.52825/ocp.v5i.1188

Published: 18 Mar. 2025

A Lightweight Approach to FDOs via Bioschemas, RO-Crate and Signposting on GitHub Pages

Rohitha Ravinder[1], Nelson Quiñones[1], Dietrich Rebholz-Schuhmann[1,2], and Leyla Jael Castro[1,*]

[1]ZB MED Information Centre for Life Sciences, Cologne, Germany

[2]University of Cologne, Cologne, Germany

*Correspondence: LJ Castro, ljgarcia@zbmed.de

Abstract. Here we present a proof-of-concept using Bioschemas, RO-Crate and Signposting as a lightweight approach to FDOs describing research outcomes exposed on GitHub pages. Research artifacts produced by a research and development team, namely the SemTec team at ZB MED, are exposed via GitHub pages and enriched with structured metadata using schema.org and Bioschemas. Research artifacts corresponding to the same research project are put together in an RO-Crate. Signposting is used on the GitHub landing pages. The combination of these three elements facilitates rich FAIR metadata for research artifacts.

Keywords: RO-Crate, Signposting, Research Metadata, FDO

1. Background

RO-crate [1] provides a lightweight approach to package research outcomes in the form of digital objects together with their corresponding metadata, using and extending schema.org [2] to describe an RO-crate package. RO-crate can be easily aligned to the Findable, Accessible, Interoperable and Reusable (FAIR) principles for data [3], software [4], workflows [5], etc., thanks to the use of RO-crate profiles. In RO-crate, a profile corresponds to a set of conventions, types and properties tailored to a specific domain and, or a community. Packaged resources can be further described according to community agreements, for instance Bioschemas. Bioschemas [6] is a collaborative effort to facilitate the use of schema.org to describe research outcomes (e.g., datasets, software) and Life Science concepts (e.g., gene, protein). On its side, Signposting [7] provides typed links on the web page headers so machines can better find and understand the underlying metadata.

RO-crates can be used as a lightweight approach [8,9] to achieve FAIR Digital Objects (FDOs), complemented by Bioschemas and Signposting, which encompass community agreements and machine-friendly approaches, both of which are key elements of FAIR and FDOs. The FDO approach [10] corresponds to a series of recommendations aiming at improving FAIRness while also extending it to cover typed operations, allowing implementation via different compliant configurations [11]. In this article, we present how we have used these technologies to describe the research outcomes of our research team via GitHub pages.

2. A lightweight approach to FDOs

The ZB MED SemTec team GitHub pages expose research projects, theses, software, vocabularies, and datasets produced by the team using Bioschemas and schema.org markup.

These pages correspond to metadata landing pages, making it easier to separate and individually identify metadata and digital objects. The actual digital objects will be published in repositories adequate to their nature (e.g., scholarly articles via journals, preprints or conference proceedings, datasets in data repositories, and software releases in Zenodo). RO-crates are used on top of research projects and theses, so they are packed together with the metadata and all the downloadable relevant research outcomes as a whole, see Figure 1.

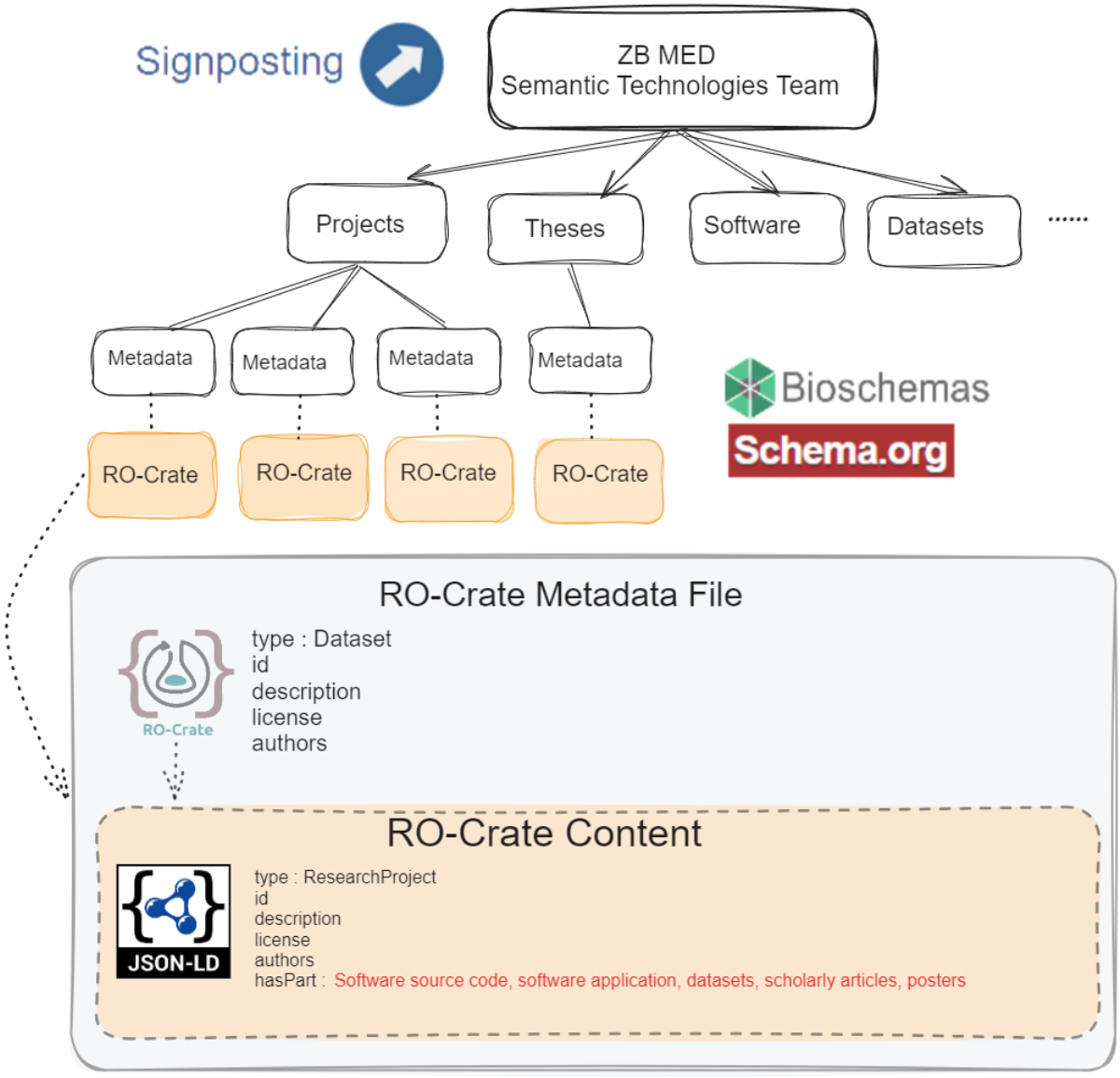

Figure 1. *Implementation of RO-crate on the ZB MED SemTec team GitHub pages. Signposting is used on the headers corresponding to RO-crate supported pages. Bioschemas and schema.org are used to describe all research outcomes (whether or not packaged as RO-crates).*

Signposting level 2 is also supported (i.e., a comprehensive set of typed links via a Link Set, making that Link Set discoverable). In this way, we are aligning to FDO configuration types 1 (directly with our metadata) and 2 (via RO-crate), see Figure 2. We use w3id.org identifiers and (semi)automation of the metadata records.

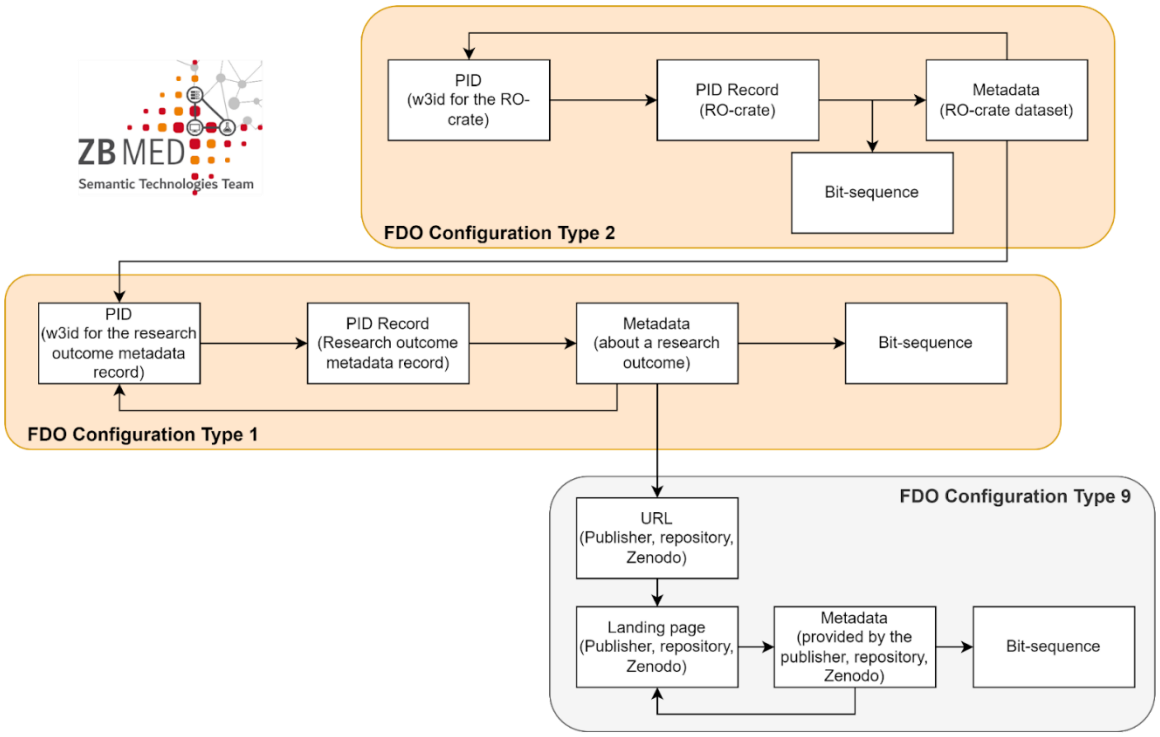

Figure 2. *FDO configuration types supported by our approach of metadata and RO-crates on GitHub pages (i.e., configuration types 1 and 2). Other configuration types (e.g., 9) could be supported by third-parties we connect to.*

3. Future work

This proof-of-concept will be later used as an approach for an FDO registry for machine learning (ML) models metadata, namely MLentory [12]; see Figure 3 for an early sketch. MLentory relies on FDOs to improve FAIrness as well as reproducibility and transparency for ML models. MLentory aggregates, harmonizes and FAIRifies data from various ML model and model-related repositories and platforms, some of them shown in Figure 3.

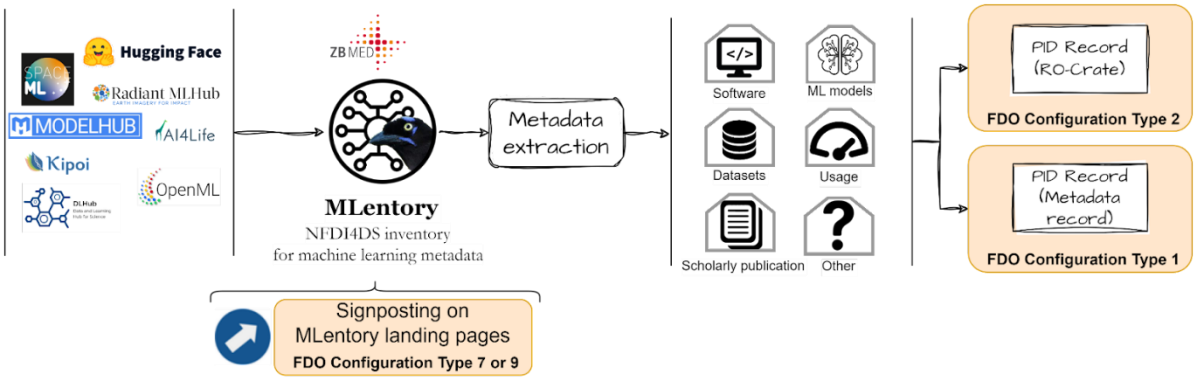

Figure 3. *Planned adoption of FDO using Signposting, RO-crate and metadata records in MLentory, an FDO registry for ML models metadata.*

Data availability statement

Metadata used on the SemTec GitHub pages can be accessed directly at the corresponding repository or harvested from the pages as it is exposed as embedded JSON-LD.

Author contributions

RR contributed to the investigation, methodology, software, writing -original draft, and writing -review and editing. NQ contributed to investigation, methodology, software, and writing -original draft. DRS contributed to the conceptualization, funding acquisition, methodology, software, supervision, and writing -original draft. LJC contributed to the conceptualization, data curation, funding acquisition, investigation, methodology, software, supervision, writing -original draft, and writing -review and editing.

Competing interests

The authors declare that they have no competing interests.

Funding

This work has been partially supported by the NFDI4DataScience project funded by the German Research Foundation (DFG) (no. 460234259), and the FAIR-Impact GA 101057344 Support Action #2: Enabling FAIR Signposting and RO-Crate for content/metadata discovery and consumption.

Acknowledgement

We acknowledge the ELIXIR BioHackathon Europe 2023, in particular project 15 “Enabling FAIR Digital Objects with RO-Crates, Signposting and Bioschemas”. We improve and finalize main aspects of our approach during the BioHackathon.

References

[1] Soiland-Reyes S, Sefton P, Crosas M, Castro LJ, Coppens F, Fernández JM, et al. Packaging research artefacts with RO-Crate. Data Science. 2022; 1–42. doi:10.3233/DS-210053
[2] Guha RV, Brickley D, Macbeth S. Schema.org: evolution of structured data on the web. Commun ACM. 2016;59: 44–51. doi:10.1145/2844544
[3] Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3: 160018. doi:10.1038/sdata.2016.18
[4] Barker M, Chue Hong NP, Katz DS, Lamprecht A-L, Martinez-Ortiz C, Psomopoulos F, et al. Introducing the FAIR Principles for research software. Sci Data. 2022;9: 622. doi:10.1038/s41597-022-01710-x
[5] Goble C, Cohen-Boulakia S, Soiland-Reyes S, Garijo D, Gil Y, Crusoe MR, et al. FAIR Computational Workflows. Data Intelligence. 2020;2: 108–121. doi:10.1162/dint_a_00033
[6] Gray AJG, Goble C, Jimenez RC. From Potato Salad to Protein Annotation. ISWC Posters and Demo session. Vienna, Austria; 2017. p. 4. Available: http://ceur-ws.org/Vol-1963/paper579.pdf
[7] Van de Sompel H. FAIR Digital Objects and FAIR Signposting. 2023 May 27. doi:10.5281/zenodo.7977333
[8] Soiland-Reyes S, Sefton P, Castro LJ, Coppens F, Garijo D, Leo S, et al. Creating lightweight FAIR Digital Objects with RO-Crate. Research Ideas and Outcomes. Pensoft Publishers; 2022. p. e93937. doi:10.3897/rio.8.e93937
[9] Castro LJ, Soiland-Reyes S, Rebholz-Schuhmann D. RO-Crates meets FAIR Digital Objects. 2023 Sep 15. doi:10.5281/zenodo.8348924

[10] De Smedt K, Koureas D, Wittenburg P. FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. Publications. 2020;8: 21. doi:10.3390/publications8020021
[11] Lannon L, Peters-von Gehlen K, Anders I, Pfeil A, Schlemmer A, Trautt Z, et al. FDO Forum FDO Configuration Types. 2022 [cited 8 Dec 2023]. doi:10.5281/zenodo.7825703
[12] Solanki D, Quiñones N, Rebholz-Schuhmann D, Castro LJ. MLentory, an FDO registry for machine learning models - Poster. Poster presented at: SWAT4HCLS 2024; PUBLISSO FRL; 2024. doi:10.4126/FRL01-006473259