1st Conference on Research Data Infrastructure

Poster presentations II

https://doi.org/10.52825/CoRDI.v1i.396

Published: 07 Sept. 2023

RO-Crates meets FAIR Digital Objects

Leyla Jael Castro[1][https://orcid.org/0000-0003-3986-0510], Stian Soliland-Reyes[2][https://orcid.org/0000-0001-9842-9718], and Dietrich Rebholz-Schuhmann[1,3][https://orcid.org/0000-0002-1018-0370]

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

2 The University of Manchester, Manchester, United Kingdom

3 University of Cologne, Cologne, Germany

Abstract. RO-Crates makes it easier to package research digital objects together with their metadata so both dependencies and context can be captured. Combined with FAIR good practices such as the use of persistent identifiers, inclusion of license, clear object provenance, and adherence to community standards, RO-crates provides a way to increase FAIRness in science. In this abstract we present the use of RO-Crates, combined with Linked Data best practices, as an implementation for lightweight FAIR Digital Objects, and its possible use in NFDI consortia.

Keywords: RO-Crates, FAIR, FAIR Digital Objects

1. Background

Linked Data (LD) [1] builds upon standards for the Web (e.g., Hypertext Transfer Protocol (HTTP), and Unique Resource Identifier (URIs)) and the Semantic Web (e.g., Resource Data Framework (RDF)), making it easier to interconnect resources to each other via meaningful links. LD can be seen as a precursor and motivator [2] of the Findable, Accessible, Interoperable and Reusable FAIR Principles [3] as it already discuss findability via URIs, accessibility using HTTP, interoperability thanks to the use of semantic data, and reusability by providing a license. Built on top of FAIR, the FAIR Digital Objects (FDOs) [4] aims at clearly separating the metadata from the digital object from its possible materializations with an additional operational layer for machines to directly act over the underlying digital object, with all these components forming a cohesive unit.

RO-Crate [5] follows best practices from LD to provide a lightweight approach to package digital objects together with their metadata, making it easier for researchers to capture both dependencies and context. RO-Crate uses structured metadata in JSON-LD and based on Schema.org [6] to define profiles corresponding to different digital objects. RO-Crate profiles encourage the use of FAIR best practices such as using Persistent Identifiers, (PIDs), providing a license, and linking to related objects. Although the use of LD and RO-Crates does not guarantee compliance to FDO specification requirements, their combination together with a set of constraints on the metadata [7] would turn RO-crates into FDOs. For instance, by requesting the declaration of the type in the metadata, assigning a separate PID to metadata, and using HTTP operations in a consistent manner.

In the rest of this abstract we introduce the use of RO-Crates to implement FDOs (building upon [8], discuss some possible uses and benefits in the National Research Data Infrastructure (NFDI) in Germany, and present some future work.

2. RO-Crates Compliance to FDO Requirements

The FDO Forum has published a set of FDO Requirement Specifications which, in their Version 1.0, comprise nine generic guidelines and twelve FDO requirements (FDORs) [9]. Here we will focus on the FDORs. Assuming an RO-crate as an FDO, we will show how RO-crates can fulfill those FDORs.

Every FDO is assigned a PID (**FDOR1**). PIDs are commonly assigned by third-parties, for instance Digital Object Identifiers (DOIs) are governed by the International DOI Foundation [10] while the Permanent Identifiers for the Web (w3id) [11] are managed by the community via Pull Requests to their GitHub repository. PIDs can be assigned to RO-crates to identify the package as a whole. **FDOR2** states that the PID assigned to an FDO resolves to an structured PID record (being a PID record an FDO, in this case an RO-Crate) following a PID Profile (defined by the community). RO-crates support the use of profiles which effectively define the metadata (types and attributes) accompanying the packaged DO. To fulfill **FDOR2**, such metadata should include the DO type. Different RO-Crate profiles correspond to different DO types for which RO-Crates rely on profiles defined by the Bioschemas [12] community (which include the DO type). Once the FDO Typing System is defined, RO-Crate and Bioschemas will need to align to those. The PID Record (here the RO-Crate) includes mandatory and optional FDO attributes and attributes agreed by the community (**FDOR3**); RO-Crates already follow community agreed types and attributes. If a bit-sequence is available for an FDO, it has to be accessible through the FDO (**FDOR4**). RO-Crates are flexible in this sense, they can contain the actual DO file, or point to a location where it is hosted. An additional effort may be required for RO-Crates to check the accessibility of files externally hosted. **FDOR5** refers to accessibility via standard protocols; currently, RO-Crates can be accessed via HTTP.

FDOR6 deals with the availability of Create, Read, Update, Delete (CRUD) operations for FDOs. RO-Crates do not directly support operations. This is an area where an extension, e.g., a supporting API, would be required. **FDOR7** aims at securing the integrity between FDO Types and operations, which should be maintained in registries. We see this as a requirement external to FDOs and more related to an FDO ecosystem providing such registries. What RO-Crates would need to do wrt **FDOR7** is making sure to keep up-to-date with the pairs type-operations maintained in such registries. For this, an external agent, e.g., RO-Crate monitor, would be needed.

RO-Crates support packaging metadata so it can be an FDO itself, **FDOR8**. RO-Crates allows metadata of different types, **FDOR9**, as far as they follow an existing profile. **FDOR10**, metadata schemas are FDOs maintained by communities, needs some additional work as Bioschemas specifications are not currently compliant to FDOs. Collections are supported by RO-Crates, **FDOR11**, but their construction is not yet part of the RO-crates. Finally, deletion mechanisms, **FDOR12**, are not yet supported by RO-Crates.

3. Possible Use and Benefits in NFDI

Being FAIR a cornerstone for NFDI consortia, FDOs would be an enhancement as the operational level should improve (semi)automatic interoperability across multiple disciplines and sorts of DOs. FDO implementation via RO-Crates provides an already known environment as it relies on LD, schema.org and Bioschemas, with some NFDIs already using schema.org (e.g., NFDI4DataScience, NFDI4Chem, NFDI4Culture, NFDI4MatWerk). As RO-Crates rely on LD, they are also compatible with Knowledge Graphs, a fundamental element in many NFDIs. In addition, RO-Crates flexibility, particularly regarding the metadata accompanying the actual DO, becomes an advantage for a multidisciplinary environment such as the NFDI consortia. RO-Crates can also be combined with authorization and authentication so private data remains protected (e.g., by pointing to a protected repository rather than including the data) while its metadata is open and public.

3. Future work

RO-Crate is a promising route towards FDOs. Although it does not yet fulfill all the FDORs, there is a growing interest in the RO-Crate community in supporting the realization of FDOs and benefit of their advantages with some ongoing efforts to that end [8].

Data availability statement

This submission is not based on data.

Author contributions

LJC: conceptualization, writing – original draft, writing – review & editing. SSR: conceptualization, project administration, writing – review & editing. DRS: conceptualization, funding acquisition, project administration, writing – review & editing.

Competing interests

The authors declare that they have no competing interests.

Funding

LJC and DRS are supported by the NFDI4DataScience, funded by the Deutsche Forschungsgemeinschaft DFG, project no. 460234259. SSR is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement numbers 823830 (BioExcel-2), 824087 (EOSC-Life) and the Horizon Europe programme under grant agreement 101046203 (BY-COVID).

References

1. C. Bizer, T. Heath, and T. Berners-Lee, 'Linked Data - The Story So Far', International Journal on Semantic Web and Information Systems (IJSWIS), vol. 5, no. 3, pp. 1–22, 2009, doi:https://doi.org/10.4018/jswis.2009081901
2. A. Hasnain and D. Rebholz-Schuhmann, 'Assessing FAIR Data Principles Against the 5-Star Open Data Principles', in The Semantic Web: ESWC 2018 Satellite Events, A. Gangemi, A. L. Gentile, A. G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J. Z. Pan, and M. Alam, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, pp. 469–477. doi:https://doi.org/10.1007/978-3-319-98192-5_60
3. M. D. Wilkinson et al., 'The FAIR Guiding Principles for scientific data management and stewardship', Sci Data, vol. 3, no. 1, Art. no. 1, Dec. 2016, doi:https://doi.org/10.1038/sdata.2016.18
4. K. De Smedt, D. Koureas, and P. Wittenburg, 'FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units', Publications, vol. 8, no. 2, Art. no. 2, Jun. 2020, doi:https://10.3390/publications8020021
5. S. Soiland-Reyes et al., 'Packaging research artefacts with RO-Crate', Data Science, pp. 1–42, Jan. 2022, doi:https://doi.org/10.3233/DS-210053
6. R. V. Guha, D. Brickley, and S. Macbeth, 'Schema.org: evolution of structured data on the web', Commun. ACM, vol. 59, no. 2, pp. 44–51, Jan. 2016, doi:https://doi.org/10.1145/2844544 .

7. S. Soiland-Reyes, L. J. Castro, D. Garijo, M. Portier, C. Goble, and P. Groth, ‘Updating Linked Data practices for FAIR Digital Object principles’, in Research Ideas and Outcomes, Pensoft Publishers, Oct. 2022, p. e94501. doi:https://doi.org/10.3897/rio.8.e94501
8. S. Soiland-Reyes et al., ‘Creating lightweight FAIR Digital Objects with RO-Crate’, in Research Ideas and Outcomes, Pensoft Publishers, Oct. 2022, p. e93937. doi:https://doi.org/10.3897/rio.8.e93937
9. G. Strawn, P. Wittenburg Eds. “FDO Forum FDO Requirement Specifications Version 1.0”. Available at https://docs.google.com/document/d/1-4_yGRrIcgdMIwaFvHyUt6lxDdfzGpqLUCEihE0vJ-g (accessed 2023.04.26)
10. DOI Foundation. “DOI Foundation”. https://www.doi.org/ (accessed 2023.04.26)
11. W3ID Community. “Permanent Identifiers for the Web”. https://w3id.org/ (accessed 2023.04.26)
12. A. J. G. Gray, C. Goble, and R. C. Jimenez, ‘From Potato Salad to Protein Annotation’, in ISWC Posters and Demo session, Vienna, Austria, Oct. 2017, p. 4. [Online]. Available: http://ceur-ws.org/Vol-1963/paper579.pdf