

Placeholder Name of the conference

Placeholder Session title

<https://doi.org/10.52825/xxxx.....> DOI placeholder (WILL BE FILLED IN BY TIB Open Publishing)

© Authors. This work is licensed under a Creative Commons Attribution 4.0 International License

Published: (WILL BE FILLED IN BY TIB Open Publishing)

Enriching Webby FDOs with Provenance in MLentory

Rohitha Ravinder¹  <https://orcid.org/0009-0004-4484-6283>, Leyla Jael Castro^{1,*}  <https://orcid.org/0000-0003-3986-0510>,
and Dietrich Rebholz-Schuhmann^{1,2}  <https://orcid.org/0000-0002-1018-0370>

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

²University of Cologne, Cologne, Germany

*Correspondence: Leyla Jael Castro, ljgarcia@zbmed.de

Abstract. Ensuring FAIRness for digital research artifacts requires not only structured metadata but also explicit provenance. MLentory, a registry for machine learning model metadata, already implements the Webby FAIR Digital Object (FDO) approach, exposing model metadata through RO-Crates and FAIR Signposting. Building on this existing infrastructure, we plan to extend MLentory with a lightweight provenance layer to capture how metadata is generated and enriched as well as tracing its original sources. This planned enhancement will enable machine-actionable tracking of metadata lineage, improving transparency, traceability and interoperability.

Keywords: Webby FDO, Signposting, RO-Crate, PROV-O, Metadata

1. Background

MLentory [1] is a registry for machine learning (ML) models developed within the NFDI4DataScience initiative. It currently aggregates, harmonizes and exposes metadata about ML models harvested from various sources via an Extraction, Transformation and Loading (ETL) pipeline. The metadata, rather than the models themselves, constitutes the primary research object, including information on technical characteristics, datasets and related publications.

Within the existing Webby FDO framework [2], MLentory provides FAIR Signposting [3] on each model's landing page and a dynamically generated RO-Crate [4] containing structured metadata and related resources. Each RO-Crate is assigned a persistent identifier (PID) via w3id.org [5], complying with FDO Configuration Type 2 [6], ensuring machine-actionable metadata access and interoperability [7].

As a next step, we plan to extend MLentory's infrastructure with a formal provenance layer. This extension will capture how RO-Crates and metadata are generated, trace their original sources and represent derivation chains. In particular, we aim to record the activities performed during metadata extraction and enrichment, identify the primary data sources harvested by MLentory and link to metadata provided by these sources. The implementation will use a combination of PROV-O [8], Dublin Core Terms [9], and schema.org [10], as it best fits our needs. Our poster submission will demonstrate this approach and outline how minimal semantic extensions can enrich the provenance of FAIR digital objects.

Data availability statement

This submission is not based on data as it corresponds to an exploration of different provenance approaches.

Author contributions

LJC and DRS have contributed to conceptualization, supervision, and funding acquisition. LJC has contributed to writing –review & editing. RR has contributed to writing –original draft and review & editing.

Competing interests

The authors declare that they have no competing interests.

Funding

This work has been partially supported by the NFDI4DataScience consortium funded by the German Research Foundation (DFG –Deutsche Forschungsgemeinschaft), project number [460234259](#).

References

- [1] Quiñones, N., Geist, L., Ravinder, R., Solanki, D., Venkatesh, S., Rebholz-Schuhmann, D., & Castro, L. J. (2025, August 4). MLentory: NFDI4DS registry for machine learning models and related artifacts. CoRDI 2025. <https://doi.org/10.5281/zenodo.16735316>
- [2] Soiland-Reyes, S., Sefton, P., Leo, S., Castro, L. J., Weiland, C., & Sompel, H. V. de. (2024). Practical webby FDOs With RO-Crate and FAIR Signposting: Experiences and Lessons Learned. Open Conference Proceedings, 5. <https://doi.org/10.52825/ocp.v5i.1273>
- [3] Van de Sompel, H. (2023, May 27). FAIR Digital Objects and FAIR Signposting. <https://doi.org/10.5281/zenodo.7977333>
- [4] Soiland-Reyes, S., Sefton, P., Crosas, M., Castro, L. J., Coppens, F., Fernández, J. M., Garijo, D., Grüning, B., La Rosa, M., Leo, S., Ó Carragáin, E., Portier, M., Trisovic, A., RO-Crate Community, Groth, P., & Goble, C. (2022). Packaging research artefacts with RO-Crate. *Data Science*, 1–42. <https://doi.org/10.3233/DS-210053>
- [5] w3id.org—Permanent Identifiers for the Web. (n.d.). Retrieved October 31, 2025, from <https://w3id.org/>
- [6] Lannon, L., Peters-von Gehlen, K., Anders, I., Pfeil, A., Schlemmer, A., Trautt, Z., & Wittenburg, P. (2022). FDO Forum FDO Configuration Types. <https://doi.org/10.5281/zenodo.7825703>
- [7] Ravinder, R., Quiñones, N., Rebholz-Schuhmann, D., & Castro, L. J. (2025). Implementing Webby FAIR Digital Objects in MLentory. 1207. <https://dl.gi.de/handle/20.500.12116/47528>
- [8] Lebo, T., Sahoo, S., & McGuinness, D. (n.d.). PROV-O: The PROV Ontology (Versions 2013-04-30). Retrieved October 31, 2025, from <https://www.w3.org/TR/prov-o/>
- [9] Dublin Core. (2020). DCMI Metadata Terms. <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

- [10] Guha, R. V., Brickley, D., & Macbeth, S. (2016). Schema.org: Evolution of structured data on the web. *Communications of the ACM*, 59(2), 44–51. <https://doi.org/10.1145/2844544>