3rd FAIR Digital Objects 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)

Comparing FDO Flavors

Transformation of a Conventional Data Monolith with Ad-Hoc Semantics into Three FDO Networks Based on Nanopublications, Handles, and RO-Crates/Signposting

Christian Backe[1,*] https://orcid.org/0009-0002-7114-0687,
and Veit Briken[1] https://orcid.org/0009-0001-8153-237X

[1]Deutsches Forschungszentrum für Künstliche Intelligenz, Germany
*Correspondence: Christian Backe, christian.backe@dfki.de

Extended Abstract. The FAIR data community in general and the FDO community in particular lack practical examples of successful implementations of the FAIR principles and of the FDO specifications. Further, data producers who wish to adopt FDOs for their data modeling and provision can choose between different FDO flavors, including Nanopublications, Handle-based FDOs, and RO-Crates with Signposting. Are these flavors mutually compatible? Do they have equivalent features? Are they equally usable?

This talk presents the transformation of a conventional data corpus into linked FDO sets of all three flavors mentioned above. It is intended as a best practise example for the implementation of the FAIR principles and the FDO specifications, and it shall allow a comparison between the different FDO flavors on the basis of concrete FDO instances. The data was acquired and used in the domain of experimental robotics field research, but the data modeling approaches and conclusions presentend in this talk are intended to be applicable to any research domain and compatible with the entire FDO ecosystem.

The original data content includes high-resolution sensor measurements from several modalities, field logbooks that contain structured experiment documentation, and specifications that provide metadata and context about data structures and used equipment. In order to comply with the FAIR principles, the corpus was made publically available on Zenodo [1]. This initial version features a clear layout, open formats, and rich semantics to facilitate reuse.

But the initial version has two major shortcomings which motivated the FDO transformation.

Firstly, the initial corpus version is designed as a monolith, i.e., as a highly integrated unit, melding 250 million datapoints into 3 Zip archives. This compromises its reusability with regard to, e.g., selective access, possible source mutation, or provenance of derived results. By transforming the monolith into a network of FDOs, data elements become addressable at a granular scale which allows, e.g., precise linking of propositions to specific observations, accurate error reporting and versioning, and flexible recombination of dataset components.

Secondly, the semantics of the initial corpus version have an ad-hoc design. They do not employ standardized terminology; semantic information is fragmented and encoded in various formats, including an SQL schema, different kinds of JSON assets, and Markdown text. The FDO transformation establishes controlled semantics uniformly across all levels of the data model, aligned with established terminologies and ontologies.

The FDO transformation was started with Semantic Web technologies (RDF, OWL, SHACL) on the Nanopublication infrastructure [2] as a use case of the FDO Connect project [3]. Due to the short project runtime of six months, the transformation first concentrated on some essential features: definition of typical FDO records representing experiment logbook payload units (CSV table rows) and specifications of their property items (key-value-pairs) [4]; definition of essential FDO profiles to bootstrap the recursive FDO profile validation chain and to demonstrate two extreme ends of potential FDO profile specificity [5]; introduction of a domain-specific foundational ontology integrated with the corpus specification and aligned with established ontologies to ensure semantic interoperability with the outside world [6]. Overall, the initial FDO data model was designed with a "depth-first" approach, sacrificing breadth due to time constraints.

The further development of the FDO transformation is currently ongoing in multiple directions: The Nanopublication-based implementation of FDO profiles and records is being replicated with Handle-based FDOs and with RO-Crates/Signposting. Additional FDO profiles are being designed to further explore the profile specificity spectrum, including bespoke corpus-specific profiles and more generic profiles for community reuse. Additional FDO records are being designed to achieve full corpus coverage regarding payload data (high-resolution sensor measurements) and metadata specifications (experiment parameters, hardware, software). Results and conclusions of these developments will be presented in this talk.

This presentation is a case study for modular multi-layered data modeling, and a practical implementation of the FDO specifications. Three FDO flavors (Nanopublications, Handle-based FDOs, RO-Crates/Signposting) are compared on the basis of concrete FDO instances with regard to flavor compatibility, feature parity, and usability from a data production perspective.

Keywords: Nanopublications, Handle-based FDOs, RO-Crates/Signposting, Experimental Research, Robotics

Data availability statement

The data can be accessed online at the locations specified in references [1], [4], [5], and [6].

Author contributions

Christian Backe: Conceptualization, Data curation, Investigation, Methodology, Software, Writing - original draft, Writing - review & editing

Veit Briken: Funding acquisition, Project administration, Supervision, Validation, Writing - review & editing

Competing interests

The authors declare that they have no competing interests.

Funding

This work is supported by the project FDO Connect, sponsored by MISSION KI — National Initiative for Artificial Intelligence and Data Economy, a project of acatech — National Academy of Science and Engineering, funded by the Federal Ministry for Digital and Transport Affairs (BMDV).

This work is supported by the project RoBivaL, funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) under grant number 50RP2150.

The authors would like to thank the Federal Government and the Heads of Government of the Länder, as well as the Joint Science Conference (GWK), for their initiative within the framework of the NFDI4Ing consortium (German Research Foundation (DFG) - project number 442146713).

The dataset has also been supported by the ZLA (NiMWK, Volkswagenstiftung, ZDIN 11-76251-14-3/19) and the Experimentierfeld Agro-Nordwest (BMEL 28DE103B18).

References

[1] C. Backe et al., "RoBivaL data corpus", Dataset, Zenodo, 2024, doi: https://doi.org/10.5281/zenodo.8424932

[2] knowledge pixels. "Nanodash Home". nanodash. https://nanodash.knowledgepixels.com (2025-11-02)

[3] DFKI. "FDO Connect". German Research Center for Artificial Intelligence GmbH, Robotics Innovation Center. https://robotik.dfki-bremen.de/en/research/projects/fdo-connect (2025-11-02)

[4] C. Backe et al. "RoBivaL FDORecord", Dataset, w3id.org / GitHub, 2025, https://w3id.org/RoBivaL/FDORecord (2025-11-02), https://github.com/cbacke/RoBivaL/tree/main/FDORecord (2025-11-02)

[5] C. Backe et al. "RoBivaL FDOProfile", Dataset, w3id.org / GitHub, 2025, https://w3id.org/RoBivaL/FDOProfile (2025-11-02), https://github.com/cbacke/RoBivaL/tree/main/FDOProfile (2025-11-02)

[6] C. Backe et al. "RoBivaL FDOProfile", Dataset, w3id.org / GitHub, 2025, https://w3id.org/ExperimentalResearchOntology (2025-11-02), https://raw.githubusercontent.com/cbacke/ExperimentalResearchOntology/refs/heads/main/ero.ttl (2025-11-02)