



## RO-Crate community update 2024

### Document Version

Submitted manuscript

[Link to publication record in Manchester Research Explorer](#)

### Citation for published version (APA):

RO-Crate community, Soiland-Reyes, S., Sefton, P., Leo, S., Castro, L. J., & Weiland, C. (2024). RO-Crate community update 2024. Manuscript submitted for publication. In *International FAIR Digital Objects Implementation Summit 2024* TIB Open Publishing.

### Published in:

International FAIR Digital Objects Implementation Summit 2024

### Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

### General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

### Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact [uml.scholarlycommunications@manchester.ac.uk](mailto:uml.scholarlycommunications@manchester.ac.uk) providing relevant details, so we can investigate your claim.



# RO-Crate community update 2024

[Stian Soiland-Reyes](#), [Peter Sefton](#), [Simone Leo](#), [Leyla Jael Castro](#), [Claus Weiland](#), [RO-Crate Community](#)

*Here we give an update of the community development and adoption of RO-Crate for FAIR Digital Object since FDO2022. It is notable that programmatic access and more detailed profiles have received high attention, as well as several FDO implementations that use RO-Crate.*

## RO-Crate community

The [RO-Crate community](#) supports and promotes FAIRification of research outcomes and provides tools for researchers to improve FAIRness. RO-Crate [[Soiland-Reyes 2022](#)] makes it easy to package research outcomes in the form of digital objects together with their corresponding metadata, using and extending [schema.org](#) to describe an RO-crate package as a dataset and all its elements (i.e., research outcomes). Thanks to the use of [profiles](#), researchers can choose a description that suits best to their needs. A [Profile Crate](#) provides a specification including types and properties tailored to a specific domain and or community.

## Workflows and provenance

Workflow Run RO-Crate [[Leo 2023](#)] is a set of three RO-Crate profiles to capture the provenance of computational workflow execution: [Process Run Crate](#), which deals with possibly composite computations not necessarily described by a formal workflow; [Workflow Run Crate](#), which extends Process Run Crate with specifications to represent computations orchestrated by a workflow; [Provenance Run Crate](#), which extends Workflow Run Crate with guidelines on how to represent the execution of the various tools used by the workflow. The profiles are developed by a working group that includes workflow developers and users, workflow engine developers and users and other members interested in the provenance of workflow execution. Workflow Run RO-Crate is already implemented by six workflow engines: Galaxy, COMPSs, StreamFlow, WfExS-backend, Sapporo and Autosubmit, enabling interoperable comparisons between heterogeneous workflow executions.

[LivePublication](#) [[Ellerm 2023](#)] is a proof of concept of an *executable paper*, which interactive visualization and statistical calculations can be regenerated on the fly taking into consideration data sources updated after the paper's publication date. Here RO-Crate enables execution on the Globus infrastructure through an innovative use of individual RO-Crates and containers for each computable element of the paper, nested within a top-level Crate for the paper. This novel approach shows how it is possible to use RO-Crate as a machine-actionable object, which do not rely on bundling an underlying workflow representation in an existing workflow language.

## TRE-FX and TREs

Trusted Research Environments (TREs) are used to provide computational analysis access to sensitive data. An architecture for federated analytics running workflows across such TREs [[Giles 2023](#)] developed the [Five Safes Crate](#) profile [[Soiland-Reyes 2023](#)], where the crate goes through manual and automatic review processes which are tracked within the crate and abstracts the workflow run requests so the federated analytics APIs are interoperable across different workflow engines.

## RO-Crate and FDO

RO-Crate has previously been proposed as a mechanism for implementing FDO [[Castro 2023](#)], in particular for metadata FDOs. The FAIR Digital Object approach fosters self-contained and machine actionable knowledge units which persistently bind necessary contextual metadata and typed operation semantics to enable processing of their actual content across different data spaces

[[Wittenburg 2023](#)]. In addition to the common development plan for FDOs leveraging i.a. on components such as Handles acting as PIDs resolving to an FDO Record and the Digital Object Interface Protocol (DOIP) for interacting with FDOs [[Blanchi 2023](#), [DONA 2018](#)], an implementation path for web-based or “webby” FDOs has been proposed building on common web technologies involving on top of RO-Crates the provision of structured metadata with schema.org and its extension [Bioschemas](#) [[Gray 2017](#)], and the modelling of typed relationships (links) between resources with [FAIR Signposting](#) [[Soiland-Reyes 2024](#)].

### Distributed FDO architectures

The [DeSci Nodes](#) system has developed a concept *Distributed Persistent Identifier* ([dPID](#)) which act as an overlay of Interplanetary File System (IPFS). APIs for the DeSci use detached RO-Crates with dPID references. This is a novel FAIR Digital Object implementation that challenges both the traditional centralised FDO approach using the Handle system, as well as the mostly Web-based RO-Crate ecosystem.

### Using the Crate-O editor and building ad-hoc vocabularies

The [Crate-O](#) tool has been developed by Language Data Commons of Australia ([LDaCA](#)) as a general-purpose RO-Crate editor as a successor or alternative to [Describo](#). This Chrome browser-based tool can describe any local folder with resources from the Web as an RO-Crate, supporting any schema.org type and property, pluggable with other rdfs vocabularies. A mode file selection indicates recommended and required properties; mode files combine schema information on classes, properties and defined terms, with RO-Crate profile rules on how they may be defined. Notably this tool is also intended for creation of such “ad-hoc” vocabularies without need of Semantic Web, and is effectively a lightweight user interface for building [Profile Crates](#) using “Schema.org style Schemas” (SOSs). There are a number of [tools](#) associated with Crate-O which can create HTML documentation and Crate-O mode files from SOSs and infer schemas from example documents.

### LDaCA

The Language Data Commons of Australia Program ([LDaCA](#)) contributes to cultural heritage preservation by recording Australian Indigenous languages, regional languages of the Pacific, and Australian English, and making such records and annotations publicly available. LDaCA [uses RO-Crate](#) as an interchange and archive format for language data, and is providing data discovery portals and API access to data using RO-Crate-centric APIs. For instance, the [LDaCA data portal](#) uses [detached RO-crates](#) for FDO-style navigation of centralised API resources.

### Dataverse

In the research data repository software [Dataverse](#) [[Trisovic 2023](#)], [RO-Crate support](#) is implemented by multiple plugins. The [AROMA](#) (ARP RO-Crate Manager) [[Zoltán 2023](#)], used by Hungarian repository [CONCORDA](#), extends Dataverse for dynamic metadata editing using the Describo Crate Builder Web component. A [FAIR-IMPACT support action](#) in 2023 saw community development of import, [export](#) and [preview](#) plugins for RO-Crate in Dataverse, and [also](#) for the Electronic Lab Notebook format ([ELN](#), which is based on RO-Crate).

## References

- [Beier 2024] Sebastian Beier, Timo Mühlhaus, Cyril Pommier, Stuart Owen, Dominik Brillhaus, Heinrich Lukas Weil, Florian Wetzels, Gavin Chait, Daniel Arend, Manuel Feser, Gajendra Doniparthi, Jonathan Bauer, Sveinung Gundersen, Pável Vázquez (2024): **BioHackEU23 report: Enabling continuous RDM using Annotated Research Contexts with RO-Crate profiles for ISA.***BioHackrXiv* <https://doi.org/10.37044/osf.io/7y2jh>
- [Blanchi 2023] Christophe Blanchi, Maggie Hellström, Larry Lannom, Andreas Pfeil, Ulrich Schwarzmann Peter Wittenburg (2023): **Implementation of Attributes, Types, Profiles and Registries.** <https://doi.org/10.5281/zenodo.7825573>
- [Castro 2023] Leyla Jael Castro, Stian Soiland-Reyes, Dietrich Rebholz-Schuhmann (2023): **RO-Crates Meets FAIR Digital Objects.** 1st Conference on Research Data Infrastructure ([CoRDI 2023](#)), 2023-09-12/–14, Karlsruhe, Germany. *Proceedings of the Conference on Research Data Infrastructure 1* <https://doi.org/10.52825/cordi.v1i.396>
- [DONA 2018] DONA, “Digital Object Interface Protocol Specification”, [https://www.dona.net/sites/default/files/2018-11/DOIPv2Spec\\_1.pdf](https://www.dona.net/sites/default/files/2018-11/DOIPv2Spec_1.pdf)
- [Eguinoa 2023] Ignacio Eguinoa, Marek Suchánek, Vojtěch Knaisl, Jan Slifka, Paul De Geest, David López, Bjorn Gruning, Simone Leo, Stian Soiland-Reyes (2023): **BioHackEU22 Report: Enhancing Research Data Management in Galaxy and Data Stewardship Wizard by utilising RO-Crates.** *BioHackrXiv* <https://doi.org/10.37044/osf.io/24jst>
- [Ellerm 2023] Augustus Ellerm, Mark Gahegan, Benjamin Adams (2023): **LivePublication: The Science Workflow Creates and Updates the Publication** *IEEE 19th International Conference on e-Science (e-Science)* <https://doi.org/10.1109/e-Science58273.2023.10254857> (preprint not available)
- [Fouilloux 2023] Anne Fouilloux, Elisa Trasatti, Federica Foglini, Alejandro Coca-Castro, Jean Iaquina (2023): **FAIR Research Objects for realising Open Science with the EOSC project RELIANCE.** 1st International Conference on FAIR Digital Objects ([FDO 2022](#)). *Research Ideas and Outcomes* 9:e108765 <https://doi.org/10.3897/rio.9.e108765>
- [Giles 2023] Thomas Giles, Stian Soiland-Reyes, Jonathan Couldridge, Stuart Wheeler, Blaise Thomson, Jillian Beggs, Suzy Gallier, Sam Cox, Daniel Lea, Justin Biddle, Rima Doal, Naaman Tammuz, Becca Wilson, Christian Cole, Elizabeth Sapey, Simon Thompson, Professor Emily Jefferson, Phillip Quinlan, Carole Goble (2023): **TRE-FX: Delivering a federated network of trusted research environments to enable safe data analytics.** *Zenodo / DARE UK* <https://doi.org/10.5281/zenodo.10055354>
- [Gray 2017] Alasdair Gray, Carole Goble, Rafael Jimenez, Bioschemas Community (2017): **Bioschemas: From Potato Salad to Protein Annotation.** *Proceedings of the ISWC 2017 posters & demonstrations and industry tracks. CEUR Workshop Proceedings 1963* <https://ceur-ws.org/Vol-1963/paper579.pdf>
- [Nicolae 2023] Bogdan Nicolae, Tanzima Z. Islam, Robert Ross, Huub Van Dam, Kevin Assogba, Polina Shpilker, Mikhail Titov, Matteo Turilli, Tianle Wang, Ozgur O. Kilic, Shantenu Jha, Line C. Pouchard (2023): **Building the I (Interoperability) of FAIR for Performance Reproducibility of Large-Scale Composable Workflows in RECUP.** *IEEE 19th International Conference on e-Science (e-Science)* <https://doi.org/10.1109/e-science58273.2023.10254808> [preprint]

[Leo 2023] Simone Leo, Michael R. Crusoe, Laura Rodríguez-Navas, Raül Sirvent, Alexander Kanitz, Paul De Geest, Rudolf Wittner, Luca Pireddu, Daniel Garijo, José M. Fernández, Iacopo Colonnelli, Matej Gallo, Tazro Ohta, Hirotaka Suetake, Salvador Capella-Gutierrez, Renske de Wit, Bruno de Paula Kinoshita, Stian Soiland-Reyes (2023): **Recording provenance of workflow runs with RO-Crate**. *arXiv:2312.07852* <https://doi.org/10.48550/arXiv.2312.07852>

[Meurisse 2023] Marjan Meurisse, Francisco Estupiñán-Romero, Javier González-Galindo, Natalia Martínez-Lizaga, Santiago Royo-Sierra, Simon Saldner, Lorenz Dolanski-Aghamanoukjan, Alexander Degelsegger-Marquez, Stian Soiland-Reyes, Nina Van Goethem, Enrique Bernal-Delgado, On Behalf of BeYond-COVID project contributors (2023): **Federated causal inference based on real-world observational data sources: application to a SARS-CoV-2 vaccine effectiveness assessment**. *BMC Medical Research Methodology* 23:248 <https://doi.org/10.1186/s12874-023-02068-3>

[Soiland-Reyes 2022] Stian Soiland-Reyes, Peter Sefton, Mercè Crosas, Leyla Jael Castro, Frederik Coppens, José M. Fernández, Daniel Garijo, Björn Grüning, Marco La Rosa, Simone Leo, Eoghan Ó Carragáin, Marc Portier, Ana Trisovic, RO-Crate Community, Paul Groth, Carole Goble (2022): **Packaging research artefacts with RO-Crate**. *Data Science* 5(2) <https://doi.org/10.3233/DS-210053>

[Soiland-Reyes 2023] Stian Soiland-Reyes, Stuart Wheeler, Thomas Giles, Carole Goble, Philip Quinlan (2023): **TRE-FX Technical Documentation - Five Safes RO-crate**. *Zenodo* <https://doi.org/10.5281/zenodo.10376350>

[Soiland-Reyes 2024] Stian Soiland-Reyes, Leyla Jael Castro, Rohitha Ravinder, Claus Weiland, Jonas Grieb, Alexander Rogers, Christophe Blanchi, Herbert Van de Sompel (2024): **BioHackEU23 report: Enabling FAIR Digital Objects with RO-Crate, Signposting and Bioschemas**. *BioHackrXiv* <https://doi.org/10.37044/osf.io/gmk2h>

[Trisovic 2023] Ana Trisovic (2023): **Cluster Analysis of Open Research Data: A Case for Replication Metadata**. *International Journal of Digital Curation* 17(1) <https://doi.org/10.2218/ijdc.v17i1.833>

[Wittenburg 2023] Peter Wittenburg, Ulrich Schwardmann, Christophe Blanchi, Claus Weiland (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>

[Zoltán 2023] Tóth Zoltán (2023): **Az RO-Crate alapú kutatási objektum csomagolás keretrendszere az ELKH ARP platformban**. (The framework of research object packaging based on RO-Crate in the ELKH ARP platform). *NETWORKSHOP 2023*, Budapest, Hungary 2023-04-12/–14. <https://doi.org/10.31915/NWS.2023.9>