Making EOSC Research Objects FAIR with RO-Crate

A common metadata overlay for EOSC repositories

Justin Clark-Casey
RO-Crate Community
EMBL-EBI
ELIXIR Europe

Stian Soiland-Reyes
RO-Crate Community
The University of Manchester
ELIXIR UK

EOSC Symposium
Prague, CZ, 2022-11-16

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

https://doi.org/10.5281/zenodo.7323300
Is it FAIR to use these repositories?

Researchers are asked to make their research outputs FAIR – where to publish?

Thousands of public, institutional and domain-specific repositories

Help from guidance and catalogues (FAIRsharing, re3data, EOSC Catalogue)

..but how to gather and reference outputs across multiple repositories?

What about contextual information?
Aims of FAIR Research Objects

Describe and package data collections, datasets, software etc. with their metadata.
Platform-independent object exchange between repositories and services.
Support reproducibility and analysis: link data with codes and workflows.
Transfer of sensitive/large distributed datasets with persistent identifiers.
Aggregate citations and persistent identifiers.
Propagate provenance and existing metadata.
Publish and archive mixed objects and references.
Reuse existing standards, but hide their complexity.
RO-Crate: Practical and general purpose

**Infrastructure independent** – avoiding repository/service silos
*Practical, lightweight, robust*

**Familiar, developer friendly**, web native, machine- and human-readable, search-engine accessible
*Adoptable Linked Data JSON and PIDs*

**Embrace diversity, legacy, unknowns**, open-ended, multi-interpretation, self-describing, interlingua
*Adaptable Metadata Profiles*

https://www.researchobject.org/ro-crate/
Realizing FAIR Digital Objects with RO-Crate

Reference existing repositories
Re-use Web standards (JSON-LD, schema.org)
Persistent identifiers w/FAIR Signposting
Add context: people, projects, etc.

RO-Crate Metadata file
- id
- type
- description
- datePublished
- author
- organisation
- license

Structured metadata about the RO-Crate and content

RO-Crate Content
- files
- directories

By reference (PID, URL)
- https://github.com/o/script
- https://doi.org/10.5281/zenodo.5841615

https://www.researchobject.org/ro-crate/specification.html
RO-Crate is used by multiple international projects. Applied across research domains – from life sciences to cultural heritage.

https://www.researchobject.org/ro-crate/in-use/
Adding rich metadata to existing data platforms

The CS3MESH4EOSC project combines major data services into the federated ScienceMesh.

Users can collaborate across established data repositories and data science services.

FAIR Description Service (based on Describo Online) to annotate data using RO-Crate.

Domain-specific profiles for additional metadata requirements.

https://doi.org/10.5281/zenodo.7310739
https://doi.org/10.3897/rio.8.e95972
https://cs3mesh4eosc.eu/
https://arkisto-platform.github.io/tools/description/describo-online/
Data Cubes – earth observation data

The EOSC project RELIANCE use RO-Crate to package data cubes of earth observation data, along with documentation, images and workflows.

Connects to related infrastructures for execution/analysis.

Metadata includes temporal coverage, spatial coverage and vertical coverage.

ROHub publishes the archived RO-Crates to general-purpose repositories (Zenodo, B2Share) for longevity and PIDs.

Fouilloux et al (2022): International conference on FAIR Digital Objects
https://doi.org/10.24424/nz65-v565
https://doi.org/10.3897/rio.8.e93940
HMC Hub Energy: FAIR Time Series of energy consumption measurements

Capturing & describing Time series:

- Electricity
- Gas
- Heat
- Drinking Water
- Compressed air

Günther et al (2022): Use Cases in HMC – from Generation to Reuse of Data
Helmholtz Metadata Collaboration
https://doi.org/10.5281/zenodo.7157694
https://helmholtz-metadaten.de/en/fair-data-commons/overview
Annotating plant research data

The RO metadata file is automatically generated, converting from ISA folder structure and annotations. It is stored using Git LFS to support large data. Reuse established standards including ISA Model, CWL.

Investigation
Study
Assay

https://isa-tools.org/

https://nfdi4plants.de/content/learn-more/annotated-research-context.html
Building an EOSC ecosystem of FAIR Workflows

- EOSC projects BY-COVID, EOSC-Life, EuroScienceGateway, BioDT exchange rich Workflow RO-Crates within an emerging EOSC ecosystem of workflow services

- Workflow Crates transfer
  - identifiers, authors, license, workflow system
  - executable workflows in their native format (e.g. Galaxy)
  - interoperable CWL description of the workflow
  - software citations (e.g. tools used)
  - required data sources
  - test suites
  - workflow execution provenance

https://workflowhub.eu/
https://www.researchobject.org/workflow-run-crate/
The RO-Crate team is:

- Peter Selton https://orcid.org/0000-0002-3545-944X
- Stan Selland-Reyes https://orcid.org/0000-0001-9842-9718
- Eoghan O’Carraigín https://orcid.org/0000-0001-8313-2150
- Oscar Corcho https://orcid.org/0000-0002-9260-0753
- Daniel Garjo https://orcid.org/0000-0003-0454-7145
- Raul Palma https://orcid.org/0000-0003-4289-4922
- Frederik Coppens https://orcid.org/0000-0001-6563-5145
- Carole Goble https://orcid.org/0000-0003-1219-2137
- José María Fernández https://orcid.org/0000-0002-4806-5140
- Kyle Chard https://orcid.org/0000-0002-7370-4805
- Jose Manuel Gómez-Perez https://orcid.org/0000-0002-5491-6431
- Michael R Crusoe https://orcid.org/0000-0002-2961-9670
- Ignacio Eguinoa https://orcid.org/0000-0002-6190-122X
- Nick Juty https://orcid.org/0000-0002-2036-8350
- Kristi Holmes https://orcid.org/0000-0001-8420-5254
- Jason A. Clark https://orcid.org/0000-0002-3588-6257
- Salvador Capella-Gutierrez https://orcid.org/0000-0002-0309-604X
- Alasdair J. G. Gray https://orcid.org/0000-0002-5711-4872
- Stuart Owen https://orcid.org/0000-0003-2130-0865
- Alan R. Williams https://orcid.org/0000-0003-3156-2105
- Giacomo Tartari https://orcid.org/0000-0003-1130-2154
- Finn Bacal https://orcid.org/0000-0002-0048-3300
- Thomas Thelen https://orcid.org/0000-0002-1756-2128
- Hervé Ménager https://orcid.org/0000-0002-7552-1009
- Laura Rodríguez-Nava https://orcid.org/0000-0003-4929-1219
- Paul Walk https://orcid.org/0000-0003-1541-5631
- brandon whitehead https://orcid.org/0000-0002-0337-8610
- Mark Wilkinson https://orcid.org/0000-0001-6960-357X
- Paul Groth https://orcid.org/0000-0003-0183-6910
- Erich Bremer https://orcid.org/0000-0003-0223-1059
- LJ García Castro https://orcid.org/0000-0003-3986-0510
- Karl Sebby https://orcid.org/0000-0001-6022-9825
- Alexander Kanitz https://orcid.org/0000-0002-3468-0652
- Ana Trisovic https://orcid.org/0000-0003-1991-0533
- Gavin Kennedy https://orcid.org/0000-0003-3910-0474
- Mark Graves. https://orcid.org/0000-0003-3486-8193
- Jasper Koehorst https://orcid.org/0000-0001-8172-8981
- Simone Leo https://orcid.org/0000-0001-8271-5429
- Marc Portier https://orcid.org/0000-0002-5648-6848
- Paul Brack https://orcid.org/0000-0002-5432-2748
- Milan Opstersek https://orcid.org/0000-0003-1743-8300
- Bert Dreesbeke https://orcid.org/0000-0003-0522-5674
- Chenzhu Ni https://orcid.org/0000-0002-2142-1731
- Kosuke Tanabe https://orcid.org/0000-0002-9986-7223
- Tomasz Miksa https://orcid.org/0000-0002-4929-7875
- Marco La Rosa https://orcid.org/0000-0001-5383-6993
- Cedric Decruw https://orcid.org/0000-0001-6387-5988
- Andreas Czerniak https://orcid.org/0000-0003-3883-4169
- Jeremy Jay https://orcid.org/0000-0002-5761-7533
- Sergio Serra https://orcid.org/0000-0002-0792-8157
- Ronald Siebes https://orcid.org/0000-0001-8772-7904
- Shaun de Witt https://orcid.org/0000-0003-4196-3658
- Shady El Damaty https://orcid.org/0000-0002-2318-4477
- Douglas Lowe https://orcid.org/0000-0002-1248-3594
- Xuanqi Li https://orcid.org/0000-0003-1498-6205
- Seunghun Gunders https://orcid.org/0000-0001-9888-7954
- Muhammad Radifar https://orcid.org/0000-0001-9156-9478
- Rudolf Wittner https://orcid.org/0000-0002-2003-2024
- Oliver Wooland https://orcid.org/0000-0002-4565-9760
- Paul De Geest https://orcid.org/0000-0002-8940-4946
- Douglas Fils https://orcid.org/0000-0002-2257-9127
- Florian Wetzel https://orcid.org/0000-0002-5526-7138
- Raül Sirvent https://orcid.org/0000-0003-0606-2512
- Abigail Miller https://orcid.org/0000-0001-9228-2882
- Jake Emerson https://orcid.org/0000-0003-0617-9219
- Davide Fucci https://orcid.org/0000-0002-0679-4361

...you?

Join RO-Crate