# Creating lightweight FAIR Digital Objects with RO-Crate and FAIR Signposting

Stian Soiland-Reyes[1,2], Peter Sefton[3], Leyla Jael Castro[4], Frederik Coppens[5], Daniel Garijo[6], Simone Leo[7], Marc Portier[8], Paul Groth[2], Carole Goble[1]

MANCHESTER 1824
The University of Manchester

[1] The University of Manchester, UK
[2] University of Amsterdam, Netherlands
[3] The University of Queensland, Australia
[4] ZB Med, Germany
[5] VIB-UGent., Belgium
[6] Universidad Politécnica de Madrid, Spain
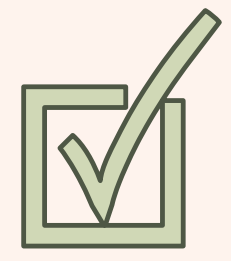[7] CRS4, Italy
[8] Vlaams Instituut voor de Zee (VLIZ), Belgium

Making FDOs practical and adoptable

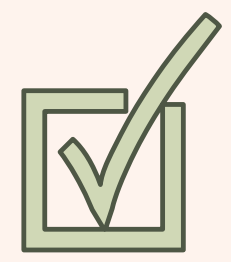**FDO envisions a distributed architecture with self-describing digital objects handling several types of metadata**
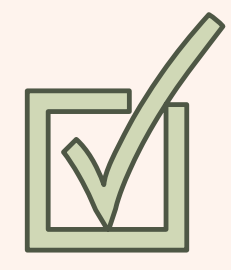
- ☑ Describe and package data collections, datasets, software etc. with their metadata
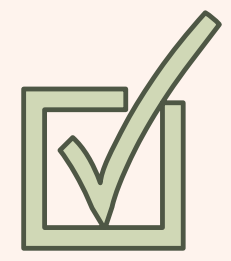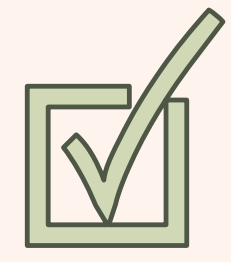
For real problems we also need…

- ☑ Platform-independent object exchange between repositories and services
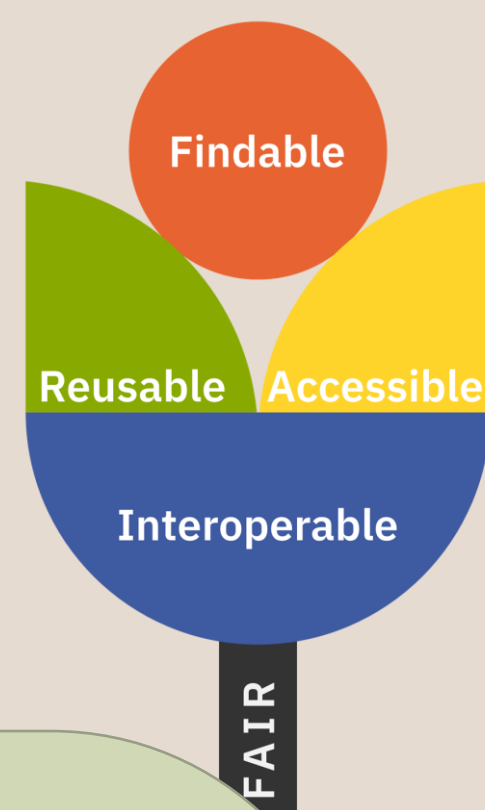- ☑ Reproducibility, linking data with codes
- ☑ Transfer of secure distributed datasets
- ☑ Citation aggregation, Provenance collection
- ☑ Mixed object publication and archiving

Findable
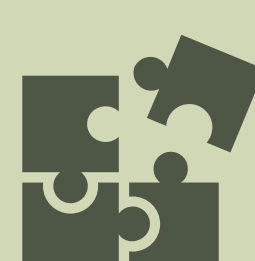Reusable Accessible
Interoperable
FAIR

RO-Crate is practical

- **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*

RO-Crate in practice

Computational Workflows
Biosciences, Climate science, Biodiversity
EOSC-Life Research Infrastructure Cluster: The computational workflow registry and its services import, export, store and publish RO-Crates, to support the full workflow life cycle. Technology is domain-agnostic and adopted outside bioscience.
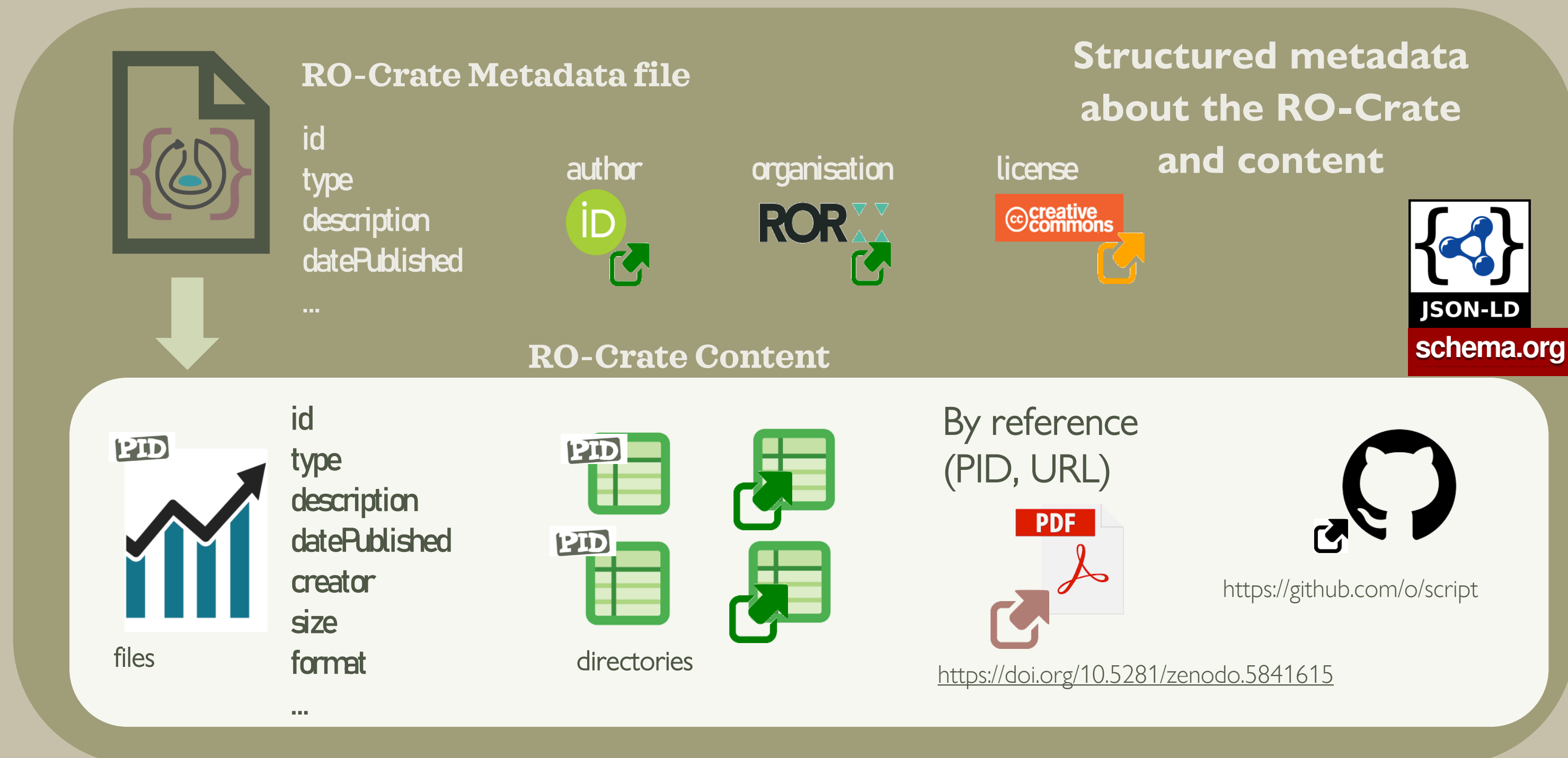
EOSC-Life   WorkflowHub

Data Cubes – *tabular data, Earth Science, Bioscience*
The EOSC project *RELIANCE* use RO-Crate to package data cubes of **earth observation data**, along with documentation, images and related infrastructures. 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.

RoHub   Reliance

Mixed Object publishing and repository exchange
The *Helmholtz HMC Hub Energy* uses RO-Crates to move time series data from different databases exported with metadata description of their structure and content into a single web service. The HERMES project uses RO-Crates for software publication pipelines

<HMC>

RO-Crate Metadata file

RO-Crate Metadata file
id
type
description
datePublished
...

author iD   organisation ROR   license creative commons

Structured metadata about the RO-Crate and content

JSON-LD schema.org

RO-Crate Content

id
type
description
datePublished
creator
size
format
...

files   directories

By reference (PID, URL)

PDF   https://github.com/o/script
https://doi.org/10.5281/zenodo.5841615

RO-Crate is a metadata FDO – simplicity, extensibility and interoperability allows data to move along with its metadata as it is transferred between applications and repositories, **independent of the FDO's storage** and access mechanisms.

RO-Crate is a **community-based** set of **best practices** for using existing **Linked Data** standards to package entities together with structured metadata.

The crate, serialized with an RO-Crate Metadata File, can also contain and describe traditional files and directories, and reference/relate external resources using their persistent identifiers.

https://doi.org/10.3233/DS-210053

PID Profile Signposting

FAIR Signposting uses standard HTTP `Link` headers (RFC8288) for any resource type (not just HTML), even if PID is unknown, so *machine agents* use HTTP `HEAD` to directly find the FDO's **PID Record**.

This *FDO PID profile* permits **any PID types** incl. **Datacite** DOI handles, **w3id** permalinks (e.g. profiles), **ORCID** (authors), stable **namespaces** (established vocabularies like schema.org), and transient **URLs** (byte downloads).

Any HTTP resource can become full FDOs by providing their own Signposting, with no other changes to their infrastructure or PIDs.

FAIR Signposting profile https://signposting.org/FAIR/
Signposting tool https://pypi.org/project/signposting/

BIODT   SYNTHESYS+   BY-COVID   FAIR-IMPACT
bioexcel   eosc EuroScienceGateway

Repository exchange and archiving
language studies and cultural heritage
PARADISEC and the *Language Data Commons of Australia* use RO-Crate for language data large text corpuses with **personally-identifiable information**. Adds **granular access control** and **restriction of use** on individual texts within the larger RO-Crate, which metadata can be open.
Jupyter Notebooks give programmatical access to crate content for analytics, selecting text by general and domain-specific metadata.

PARADISEC   LDaCA

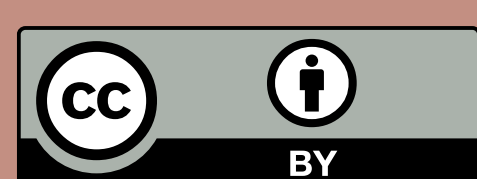### Executing Data and Software Management Plans   DSW   elixir
RO-Crates are combined with machine-actionable **Data Management Plans** (maDMPs) and Software Management Plans to automate and facilitate management of research data. Within *ELIXIR*, RO-Crate will integrate the **Data Stewardship Wizard** with Galaxy workflows to automate FDO creation that also follows data management plans.

https://researchobject.org/ro-crate
RIO abstract: https://doi.org/10.3897/rio.8.e93937
Poster: https://doi.org/10.5281/zenodo.7231713