



# Sharing data as machine-actionable objects using RO-Crate, Bioschemas and Signposting

Stian Soiland-Reyes<sup>1,2</sup>, Simone Leo<sup>3</sup>, Leyla Jael Castro<sup>4</sup>, Peter Sefton<sup>5</sup>, Carole Goble<sup>1</sup>

<sup>1</sup> The University of Manchester, UK  
<sup>2</sup> University of Amsterdam, Netherlands  
<sup>3</sup> CRS4, Italy  
<sup>4</sup> ZB Med, Germany  
<sup>5</sup> The University of Queensland, Australia

## Practical and adaptable data sharing

Building a flexible FAIR infrastructure on the Web 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...

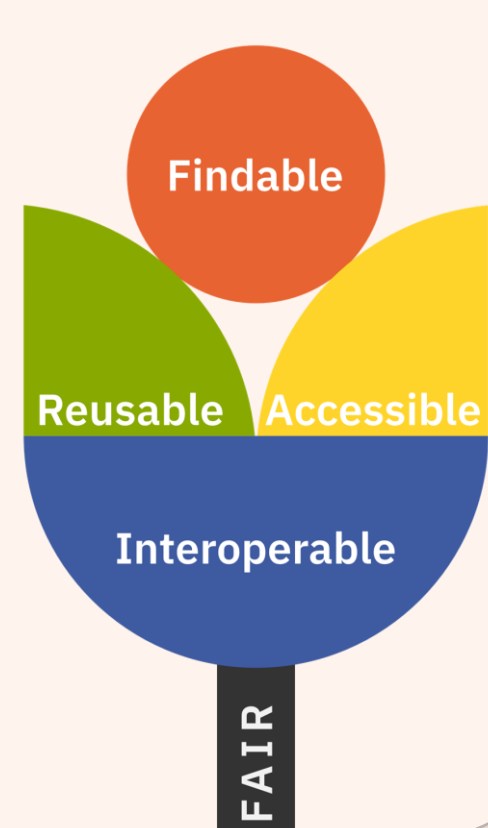
Platform-independent object exchange between repositories and services

Representation of contextual elements

Pragmatic best practices guides

Domain-specific and general-purpose profiles

Mixed object publication and archiving



## Practical approaches

**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; guidance, not restrictions*

**Embrace diversity, legacy, unknowns**, open-ended, multi-interpretation, self-describing, interlingua *Adaptable Metadata Profiles, add to existing Web resources*

## 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.

### 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.

### 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

## Bioschemas

<https://bioschemas.org/>

**Bioschemas** define domain-specific profiles to add structured metadata to Life Science resources on the Web by using and expanding **schema.org**: *ChemicalSubstance, Gene, MolecularEntity, Protein, ProteinStructure, Sample, Taxon, ...*

General-purpose profiles are being lifted for use beyond life sciences (<https://schemas.science/>) including *Dataset, Course, ComputationalWorkflow, ComputationalTool, TrainingMaterial*.

Bioschemas are deployed across >67 ELIXIR resources and beyond, covering >180 profile deployments overall.

Bioschemas are developed and maintained by working groups in an active community (>150 members)

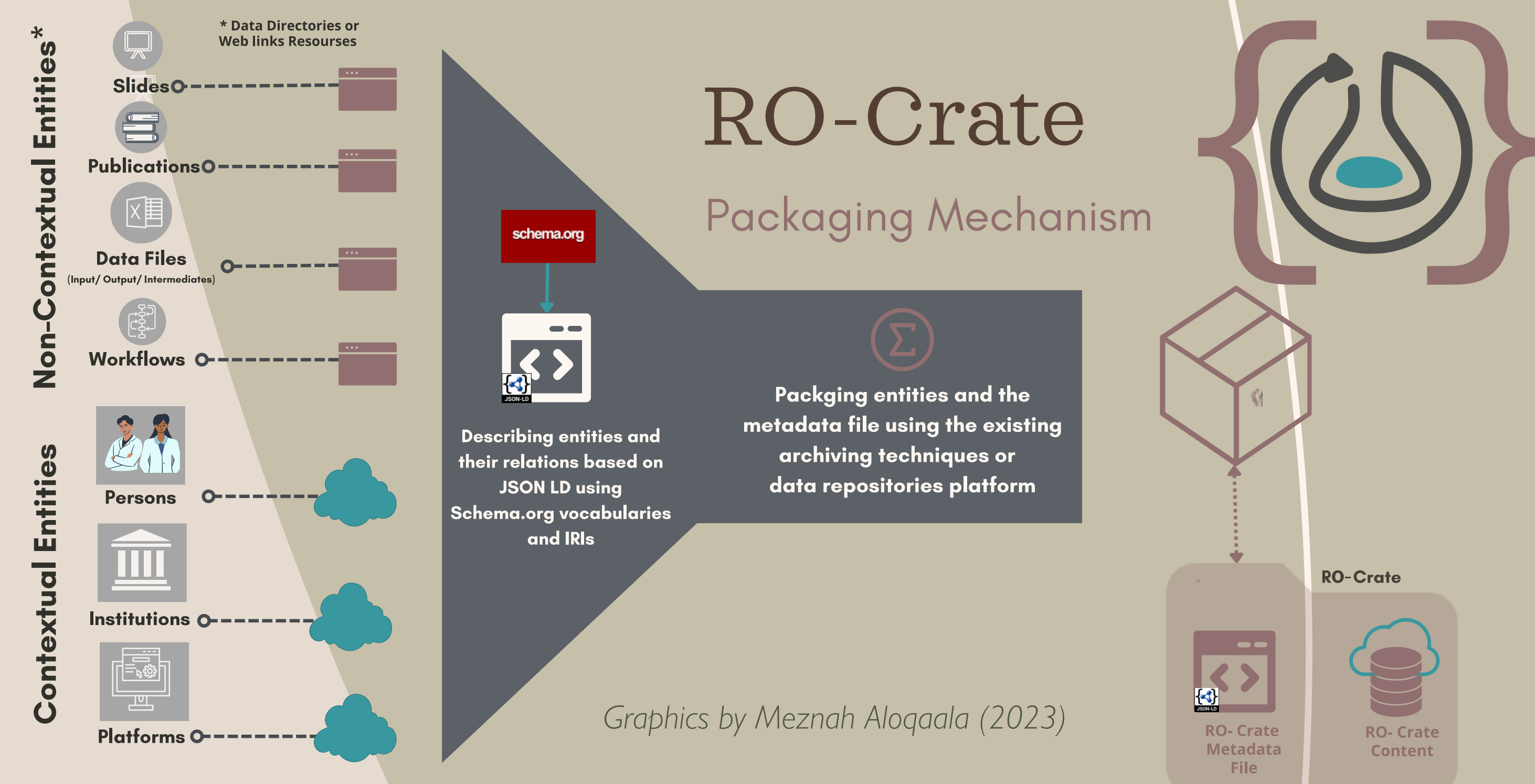
## Signposting

**FAIR Signposting** uses standard HTTP Link headers (RFC8288) for any resource type (including HTML landing pages). Machine agents then use HTTP HEAD to predictably find the individual components of a FAIR resource: *PID, authors, licence, metadata and dataset download*.

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

<https://signposting.org/FAIR/>

<https://pypi.org/project/signposting/>



Graphics by Meznah Aloqaala (2023)



### 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.

### Executing Data and Software Management Plans



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.

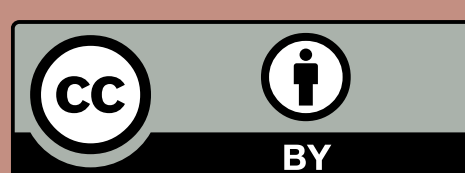
<https://bioschemas.org/>

<https://researchobject.org/ro-crate>

<https://doi.org/10.5281/zenodo.7984529>

RO-Crate and Bioschemas at ELIXIR AHM2023  
Workshop 10: Building lightweight FAIR data packages with Bioschemas and RO-Crate (Tue 11:00)  
Workshop 3: For sustainable Agriculture: Status, Collaborations and Perspectives of the Plant Community within ELIXIR (Tue 09:00)  
Mini-symposium 6: Reproducible Analytics & Infrastructure  
Workshop 17: Embedding a new data resource in the ELIXIR ecosystem – a practical guide (Wed 11:00)  
Poster 11: de-NB4-CDN – The German contribution to the Plant Community of ELIXIR  
Poster 53: From launch to maturity: growing and sustaining the RO-Mat  
Poster 87: Enabling an operational, open and FAIR EOSC ecosystem  
Poster 89: WorkflowHub – a registry for describing, sharing and publishing scientific computational workflows  
Poster 92: Improving Reproducibility of Computational Analyses Performed on the ELIXIR-GR Cloud  
Poster 93: Making workflow provenance FAIR across workflow systems with Workflow Run RO-Crate

#91



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