

Research Object Crate (RO-Crate)

An interoperable format to share research data with metadata

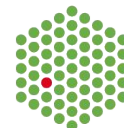
Laurent Thomas

LabID-PROV workshop - September 2026



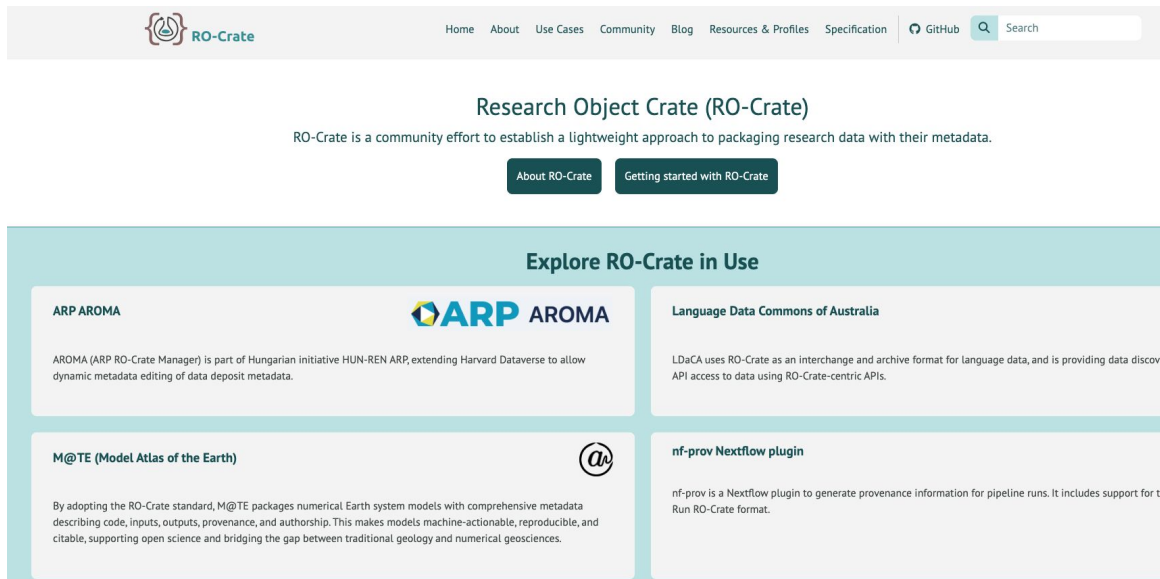
Funded by
the European Union

EMBL



RO-Crate: What and Why

- proposal to **share/publish research data** beyond simple files and folders
- **Metadata providing “context”** and explaining relations between piece of data
- used by many research communities (science to humanities)
- standardised and **machine-readable**



The screenshot shows the RO-Crate website. At the top is a navigation bar with the RO-Crate logo, links for Home, About, Use Cases, Community, Blog, Resources & Profiles, Specification, GitHub, and a search bar. Below the navigation bar, the title "Research Object Crate (RO-Crate)" is displayed, followed by a subtitle: "RO-Crate is a community effort to establish a lightweight approach to packaging research data with their metadata." Two buttons, "About RO-Crate" and "Getting started with RO-Crate", are positioned below the subtitle. The main content area is titled "Explore RO-Crate in Use" and features four cards. The first card, "ARP AROMA", includes the ARP AROMA logo and text about its role as part of the HUN-REN ARP initiative. The second card, "Language Data Commons of Australia", mentions LDaCA's use of RO-Crate for language data. The third card, "M@TE (Model Atlas of the Earth)", features the M@TE logo and describes its use of RO-Crate for Earth system models. The fourth card, "nf-prov Nextflow plugin", shows the nf-prov logo and explains its function in generating provenance information for Nextflow pipelines.

Home About Use Cases Community Blog Resources & Profiles Specification GitHub Search

Research Object Crate (RO-Crate)

RO-Crate is a community effort to establish a lightweight approach to packaging research data with their metadata.

About RO-Crate Getting started with RO-Crate

Explore RO-Crate in Use

ARP AROMA

AROMA (ARP RO-Crate Manager) is part of Hungarian initiative HUN-REN ARP, extending Harvard Dataverse to allow dynamic metadata editing of data deposit metadata.

Language Data Commons of Australia

LDaCA uses RO-Crate as an interchange and archive format for language data, and is providing data discovery API access to data using RO-Crate-centric APIs.

M@TE (Model Atlas of the Earth)

By adopting the RO-Crate standard, M@TE packages numerical Earth system models with comprehensive metadata describing code, inputs, outputs, provenance, and authorship. This makes models machine-actionable, reproducible, and citable, supporting open science and bridging the gap between traditional geology and numerical geosciences.

nf-prov Nextflow plugin

nf-prov is a Nextflow plugin to generate provenance information for pipeline runs. It includes support for Run RO-Crate format.

RO-Crates are just directories (or zip)

The ro-crate-metadata.json contains the metadata

- RO-Crates are just directories of data + **ro-crate-metadata.json**
- The json contains the metadata (a mini knowledge graph)

```
"@context": "https://w3id.org/ro/crate/1.1/context",
"@graph": [
  {
    "@id": "./",
    "@type": "Dataset",
    "author": {
      "@id": "#b2c6059a-b51b-479c-8369-4d3b03f1b4e0"
    },
    "datePublished": "2026-09-23T13:53:01+00:00",
    "description": "Workflow RO-Crate exported by LabID v26.8.0.",
    "hasPart": [
      {
        "@id": "convert.sh"
      },
      {
        "@id": "README.md"
      },
      {
        "@id": "LICENSE"
      },
      {
        "@id": "requirements.txt"
      }
    ],
    "identifier": "2982b037-d7d4-4388-aef8-15f86c436c13",
    "license": "MIT",
    "mainEntity": {
      "@id": "convert.sh"
    }
  }
]
```

- The json contains a list of **entities**
 - **Data entities** : each file has a corresponding entity
 - **Metadata entities** : an author, a software, a place

- **Entities have properties**

e.g *programming language* of a workflow file, *name* and *affiliation* of an **Author**

- Entities can be connected via **relationships** (principle of linked data)

File entity

- *format*
- *size*
- *author*

Author entity
(*name, affiliation*)

Simple “workflow RO-Crate”

bash-conversion-wf.crate.zip	
Name	
	convert.sh
	LICENSE
	README.md
	requirements.txt
	ro-crate-metadata.json

RO-Crates use ontologies for standardisation

```
"@id": "README.md",  
"@type": "File",  
"about": {  
  "@id": "./"  
},  
"encodingFormat": "text/markdown",
```

Property names corresponds to standard terms from ontologies (“controlled vocabularies”)

The list of terms is defined in the crate context (link at the top of the json
<https://www.researchobject.org/ro-crate/specification/1.1/context.jsonld>)



```
"encoding": "http://schema.org/encoding",  
"encodingFormat": "http://schema.org/encodingFormat",  
"encodingType": "http://schema.org/encodingType",
```

Ontologies ensure a common vocabulary for the metadata saved in RO-Crates

A diversity of RO-Crates

The RO-Crate profiles

- RO-Crates can be used to describe arbitrary datasets (very flexible, but very diverse)
- Base specification (minimal), freely extensible by supported specific ontology terms
- Communities have developed specialized “declinations” around specific use-cases

RO-Crate Profiles

Table of contents

1. RO-Crate Profiles

1. [RO-Crate Profiles registry](#) 

2. [Workflow RO-Crate profile](#)

1. [Workflow Testing RO-Crate profile](#) 

2. [Workflow Run Crate](#)

3. [Common Provenance Model RO-Crate profile](#)

3. [Describo profiles](#)

4. [RELIANCE RO-crate profile](#)

5. [ARC RO-Crate profile](#) 

6. [ISA RO-Crate Profile](#)

7. [ARC CWL RO-Crate Profile](#) 

8. [Electronic Lab Notebook \(ELN\)](#)

9. [RO-Crates-and-Excel](#)

10. [Making an RO-Crate profile](#)

11. [Formalizing RO-Crate profiles](#)

- Profile = custom requirements (crate MUST, SHOULD have...)
+ set of ontology terms
- **Workflow RO-Crate**
used by WorkflowHub to exchange workflows
- **Workflow Run RO-Crate**
contains run data and metadata (parameter values...)
can be generated by Nextflow, Galaxy, Snakemake

See <https://www.researchobject.org/ro-crate/profiles>

How is this useful?

RO-Crates are primarily meant as a **standard format to exchange data with arbitrarily complex metadata** (json rather “**oriented machine**” but still human readable)

Use cases

- publication to research repositories (e.g. WorkflowHub)
- interoperability between software (exporting ELN between ELN softwares)
- indexing/crawling by search engines and other algorithms

In LabID

- streamline import of workflow runs and parsing of run metadata (e.g Nextflow and Galaxy Run RO-Crate)
- publication to WorkflowHub
- export of most information available in LabID (biological and computational provenance) in one zip (soon)