

Capturing “*Just enough*” Data, Software and Metadata with RO-Crate



Stian Soiland-Reyes

eScience lab, The University of Manchester

INDElab, University of Amsterdam

@soilandreyes

<https://orcid.org/0000-0001-9842-9718>

<https://slides.com/soilandreyes/>

[10.5281/zenodo.5007432](https://doi.org/10.5281/zenodo.5007432)

@soilandreyes



FAIR Festival 2021

FAIR Implementation Challenges & Solutions

2021-06-21

H2020-INFRAEOSC-2018-2 [824087](#)

H2020-INFRAEDI-2018-1 [823830](#)

H2020-INFRAIA-2017-1 [730976](#)

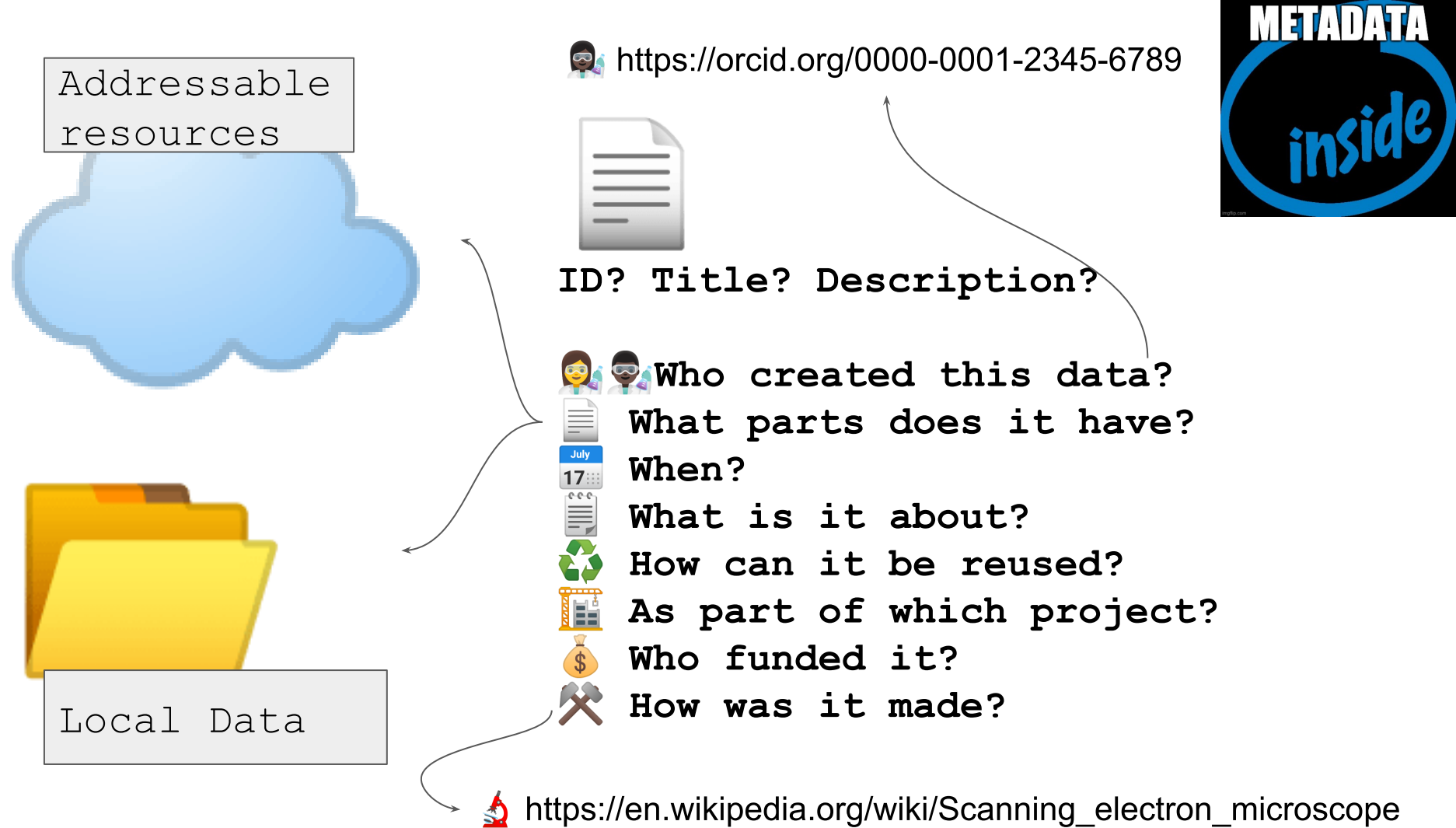
H2020-INFRADEV-2019-2 [871118](#)

H2020-INFRAIA-2018-1 [823827](#)



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

What is RO-Crate?



RO-Crate is method for self-described *datasets*
as a *digital object* using a single Linked Data
metadata document

Credit: Peter Sefton

Adapted from <https://arkisto-platform.github.io/standards/ro-crate/>



The dataset may contain *any kind* of
resource, about anything, in *any format*
as a file, URL or PID

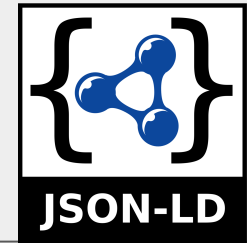
Credit: Peter Sefton

Adapted from <https://arkisto-platform.github.io/standards/ro-crate/>

```
|-- Folder1/  
|         |-- file1.this  
|         |-- file2.that  
|-- Folder2/  
|         -- file1.this  
|         -- file2.that  
|-2021-04-08 07.58.17.jpg
```



```
{  
  "@id": "2021-04-08 07.58.17.jpg",  
  "@type": "File",  
  "contentSize": 3271409,  
  "dateModified": "2021-04-08T07:58:17+10:00",  
  "description": "",  
  "encodingFormat": [  
    {"@id":  
      "https://www.nationalarchives.gov.uk/PRONOM/x-fmt/391"},  
    "image/jpeg"  
  ],  
  "name": "Cute puppy"  
},
```

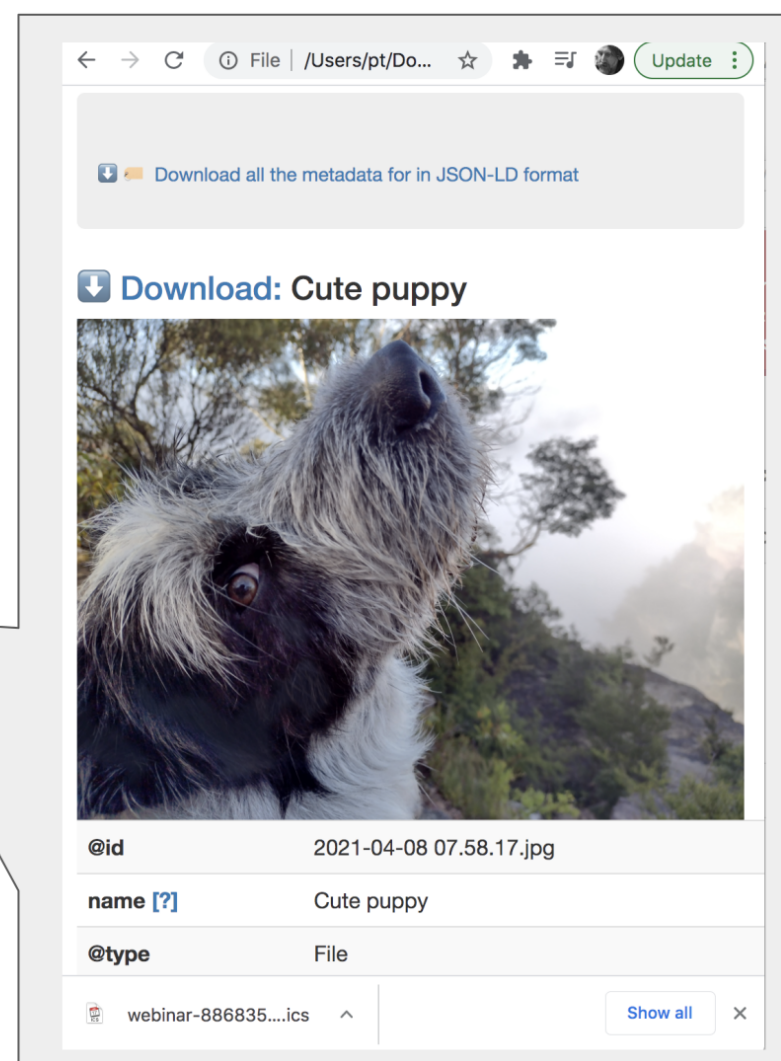


Each resource have a **machine readable**
description in JSON-LD format

Credit: Peter Sefton

Adapted from <https://arkisto-platform.github.io/standards/ro-crate/>

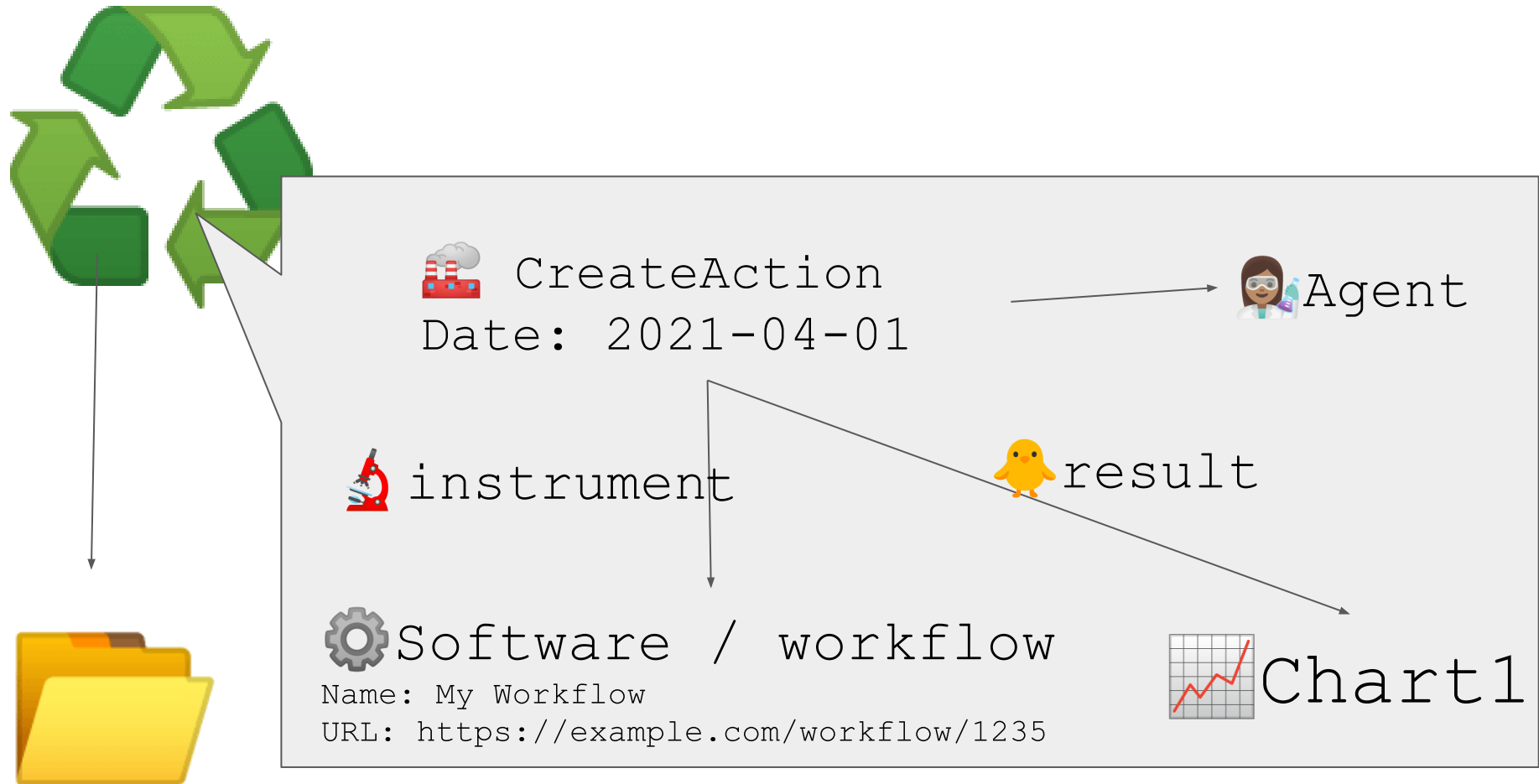
```
|-- Folder1/
|           |-- file1.this
|           |-- file2.that
|-- Folder2/
|           |-- file1.this
|           |-- file2.that
|-2021-04-08 07.58.17.jpg
```



A **human-readable** description/preview in an HTML file that *lives alongside the metadata*

Credit: Peter Sefton

Adapted from <https://arkisto-platform.github.io/standards/ro-crate/>



Provenance and workflow information can be included
– to assist in *re-use* of data and research processes

Credit: Peter Sefton

Adapted from <https://arkisto-platform.github.io/standards/ro-crate/>

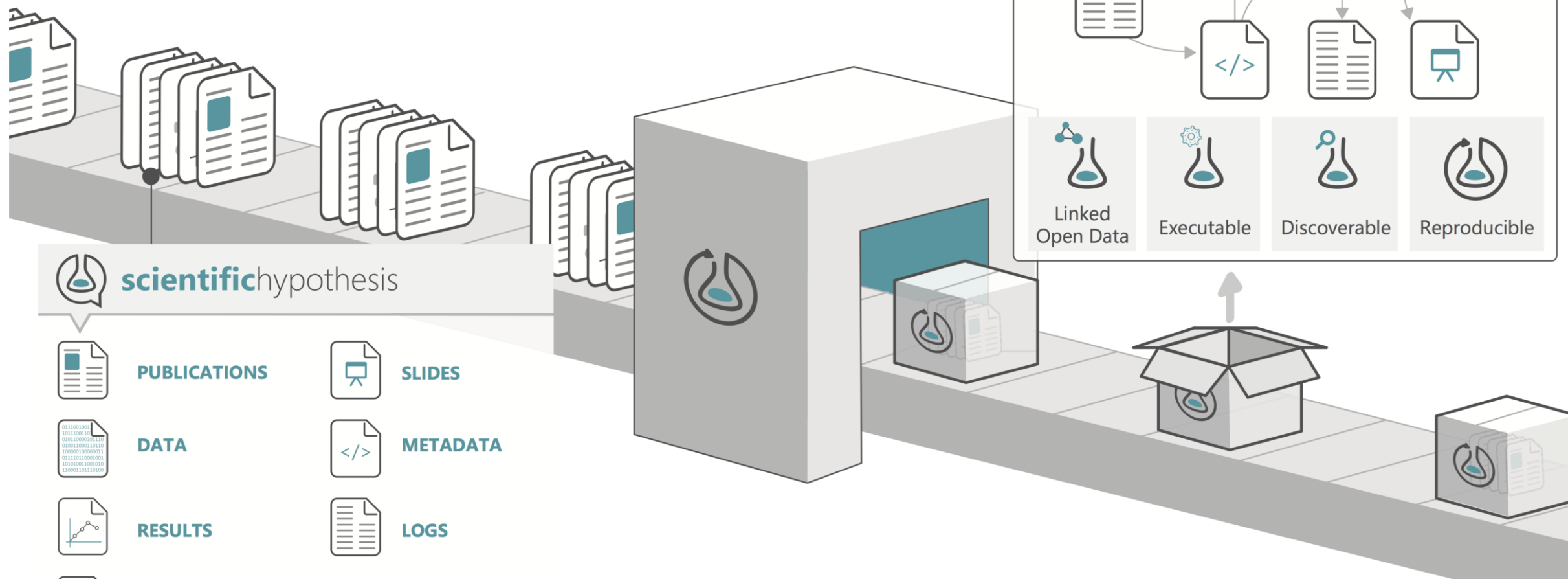


RO-Crate Digital Objects may be *packaged* for distribution eg via Zip, Bagit and OCFL
– or simply be *published on the Web*

Credit: Peter Sefton

Adapted from <https://arkisto-platform.github.io/standards/ro-crate/>

RO-Crate capture Research Objects



researchobject.org

Selected Resource: onedrive:/Sefton sample crate

Build the Collection

Manage Collection Data Files

Browse Collection Entities

Load Root Dataset

</> Add Property

name

Sample dataset for RO-Crate v1.0

citation



+ CreativeWork

+ Text



ScholarlyArticle: This is an example publication with a dodgy DOI



contactPoint



ContactPoint: Contact Peter Sefton



datePublished



+ DateTime

+ Date



June 29, 2017



description



+ Text

This is a simple dataset for demonstration purposes it contains just one image and a directory full of useless text files.



funder



+ Person

+ Organization



Organization: Australian National Data Service



Organization: University of Technology Sydney



Organization: DataCrate Project



Making your own RO-Crate with Describo

<https://arkisto-platform.github.io/describo/>[https://arkisto-
platform.github.io/tools/description/describo-](https://arkisto-platform.github.io/tools/description/describo-)*Credit: Marco La Rosa, Peter Sefton*

@soilandreyes

Dataset: Survey of Victoria Arch, Wombeyan Caves NSW

[Download this Dataset](#)

[Download all the metadata for Survey of Victoria Arch, Wombeyan Caves NSW in JSON-LD format](#)

[Check this crate](#)

<https://doi.org/10.4225/59/5a4d9b76d79f4>

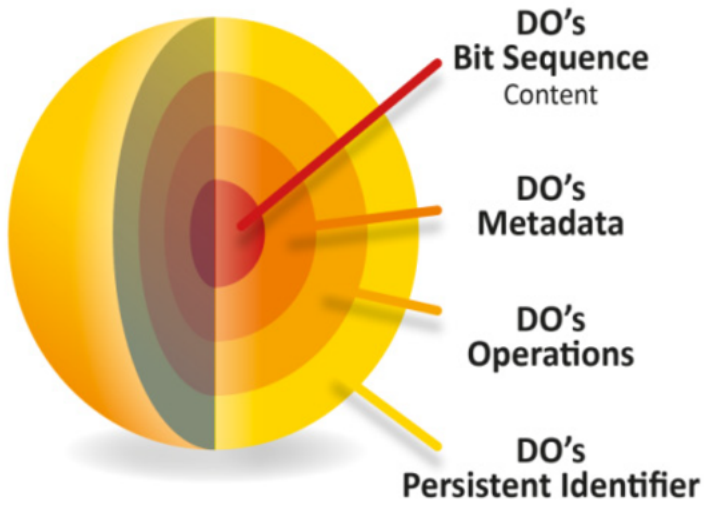
https://data.research.uts.edu.au/examples/ro-crate/0.2/Victoria_Arch_pub/

FAIR is not just machine-readable!

@id	https://dx.doi.org/10.4225/59/5a4d9b76d79f4
name	Survey of Victoria Arch, Wombeyan Caves NSW
@type	Dataset
description	This data is part of a project by Michael Lake and supported by the Australian Speleological Federation. Data was acquired at Wombeyan Caves by Robert Zlot in January 2014 using the Zebedee 3D Mapping System developed by CSIRO.
datePublished	2017-10-01
creator	<ul style="list-style-type: none">• Robert Zlot• Mike Lake• Lukas Kaul
path	./
contactPoint	Contact Mike Lake
contentLocation	Victoria Arch
hasPart	README.txt--to--wcc06_archencentre1.laz wcc04_archentrance.laz--to--README.html
identifier	<ul style="list-style-type: none">• https://dx.doi.org/10.4225/59/5a4d9b76d79f4• dx.doi.org/10.4225/59/5a4d9b76d79f4
license	Copyright CSIRO 2014: Available for research and academic purposes

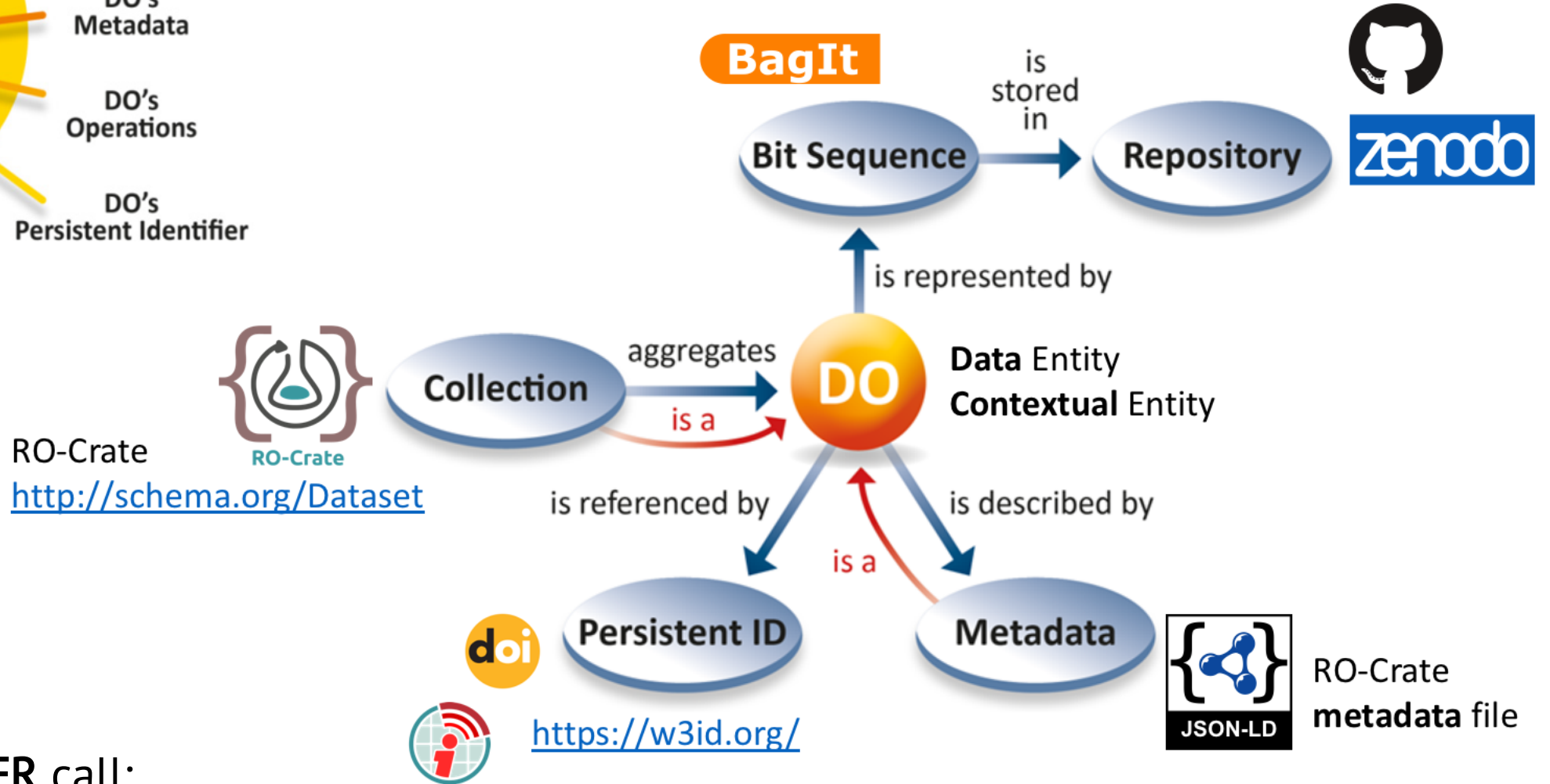
https://data.research.uts.edu.au/examples/ro-crate/0.2/Victoria_Arch_pub/ro-crate-preview.html

@soilandreyes



RO-Crate as FAIR Digital Object (FDO)

+ FAIR Signposting



2. July 2021 CWFR call:
RO-Crate relationship with FDOs
<https://osf.io/2cy86/>

@soilandreyes

Adapted from
<https://doi.org/10.3390/publications8020021>

Recordings in South Efate

Open (subject to agreeing to PDSC access conditions)

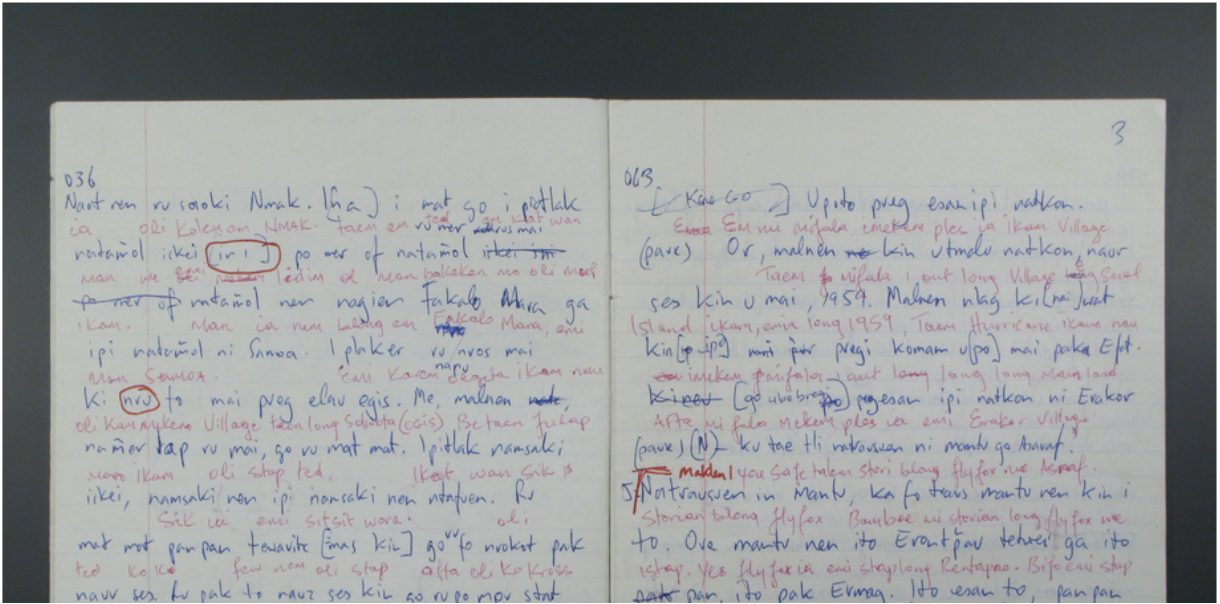
Item: NT1 / 98007
Collection: NT1
Contributors:

- Nick Thieberger (collector)
- Corin Bone (operator)
- Nick Thieberger (recorder)
- Kalsarap Namaf (speaker)
- Ikopeth (speaker)
- John Maklen (speaker)
- Waia Tenene (speaker)

NT1-98007. Text #047 (speaker is John Maklen. Text title: History of villages before Erakor); Text #048 (speaker is John Maklen. Text title: Mantu the flying fox and Erromango); Text #049. Text title: Asaraf (speaker is John Maklen);Text #050. Text title: Mumu and Kotkot (speaker is John Maklen); Text #051. Text title: Natopu ni Erakor—the spirit who lives at Erakor (speaker is John Maklen);Text #038. Text title: The need for respect (speaker is Ikopeth) Stories can be seen at NT8-TEXT. There are time-aligned transcripts of this item and handwritten transcripts by Manuel Wayne scanned as jpg files. [show less](#)



Capturing cultural heritage records as RO-Crates



ImagesAudioXML Files

NT1-98007-98007A

NT1-98007-98007A.ixt

NT1-98007-98007A.eaf

PLAY (09:22)

Natopu nen kin ito esan ga, ipi natopu nigmam.
The natopu (local spirit) that lives here, it is our natopu.

Natopu	nen	kin	ito	esan	ga,	ipi	natopu	nigmam.
natopu	nen	kin	i= to	esan	ga	i= pi	natopu	nigmam
spirit	that	COMP	3S.RS= stay	place	3S.	3S.RS= be	spirit	1P.POS

PLAY (09:26)

Ser natamō ni Erakor runomser mtaki natopu nen kin rusoso ki maarik.
Everyone from Erakor is scared of the natopu which they call 'maarik' (mister)

<https://arkisto-platform.github.io/case->

@soilandreyes



RO-Crate

RO-Crate in Earth Science

Earth Science RO-Crate profile (ongoing work)

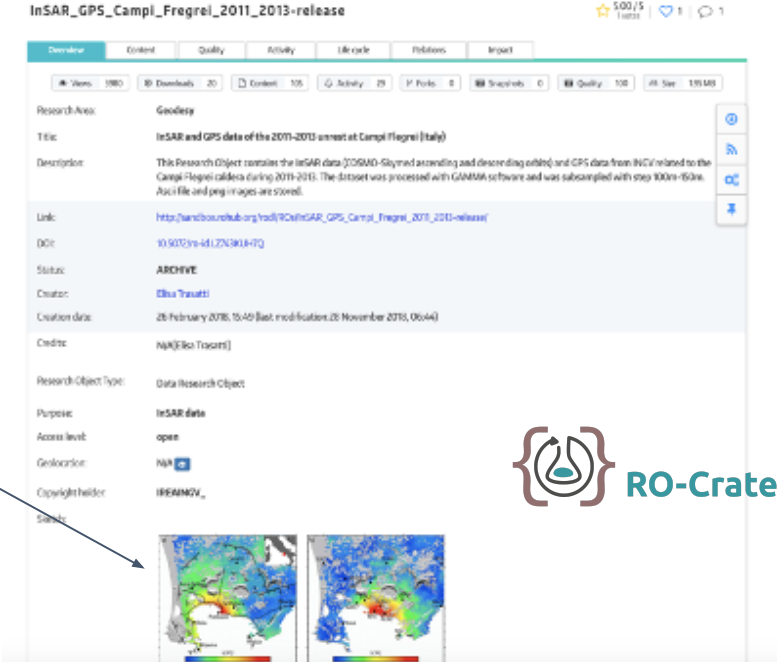
Data Cubes (**Collections and Products**) are very common in Earth Science research

We create the concept of **Data Cube Data Entity**

Relevant **metadata** for a Data Cube

- Simple **geolocation** (sch:contentLocation) and **temporal coverage** (sch:temporalCoverage)
- **Spatial coverage** (extent, resolution, crs)
- **Temporal coverage** (startTime, endTime, resolution-step)
- **Vertical coverage** (highestElevation, lowestElevation, resolution-step)
- Other (processing level, processor, instrument, platform, etc.)

http://www.rohub.org/rodetails/InSAR_GPS_Campi_Fregrei_2011_2013-release/



COMING
SOON

A valid RELIANCE RO-Crate JSON-LD graph MUST describe:

1. The [RO-Crate Metadata File Descriptor](#)
2. The [Root Data Entity](#)
3. One or more **Data Cube Data Entities**
4. Zero or more [Data Entities](#)
5. Zero or more [Contextual Entities](#)

Credit: Oscar Corcho, Carole Goble

<https://doi.org/10.5281/zenodo.4913285>

@soilandreyes

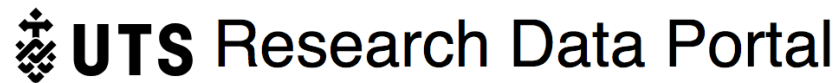
RO-Crate as a archival format

Metadata held alongside heterogeneous data

Exchange mechanism (import/export)

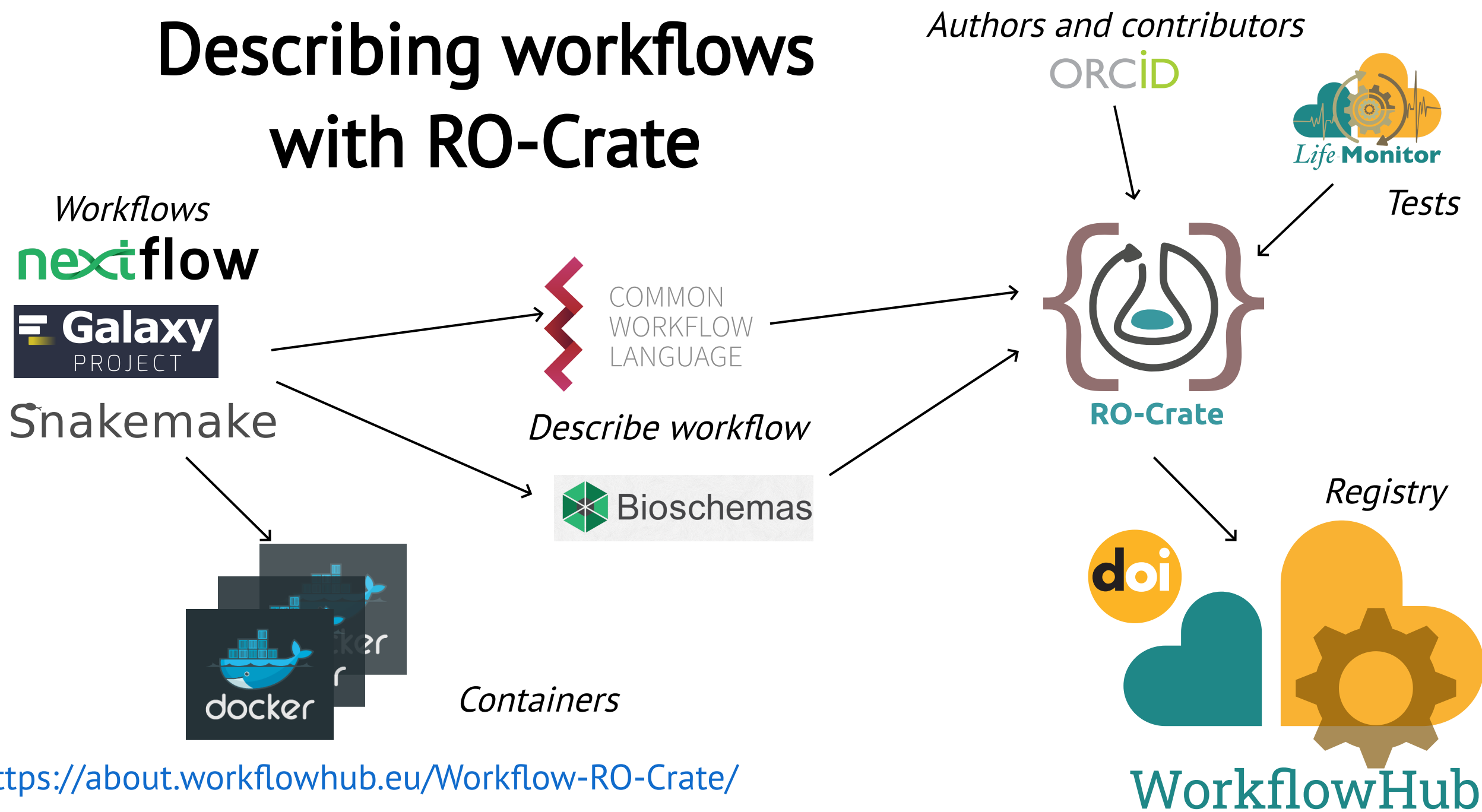
Avoid repository vendor lock-in

Extensible metadata model



Oxford Common File Layout

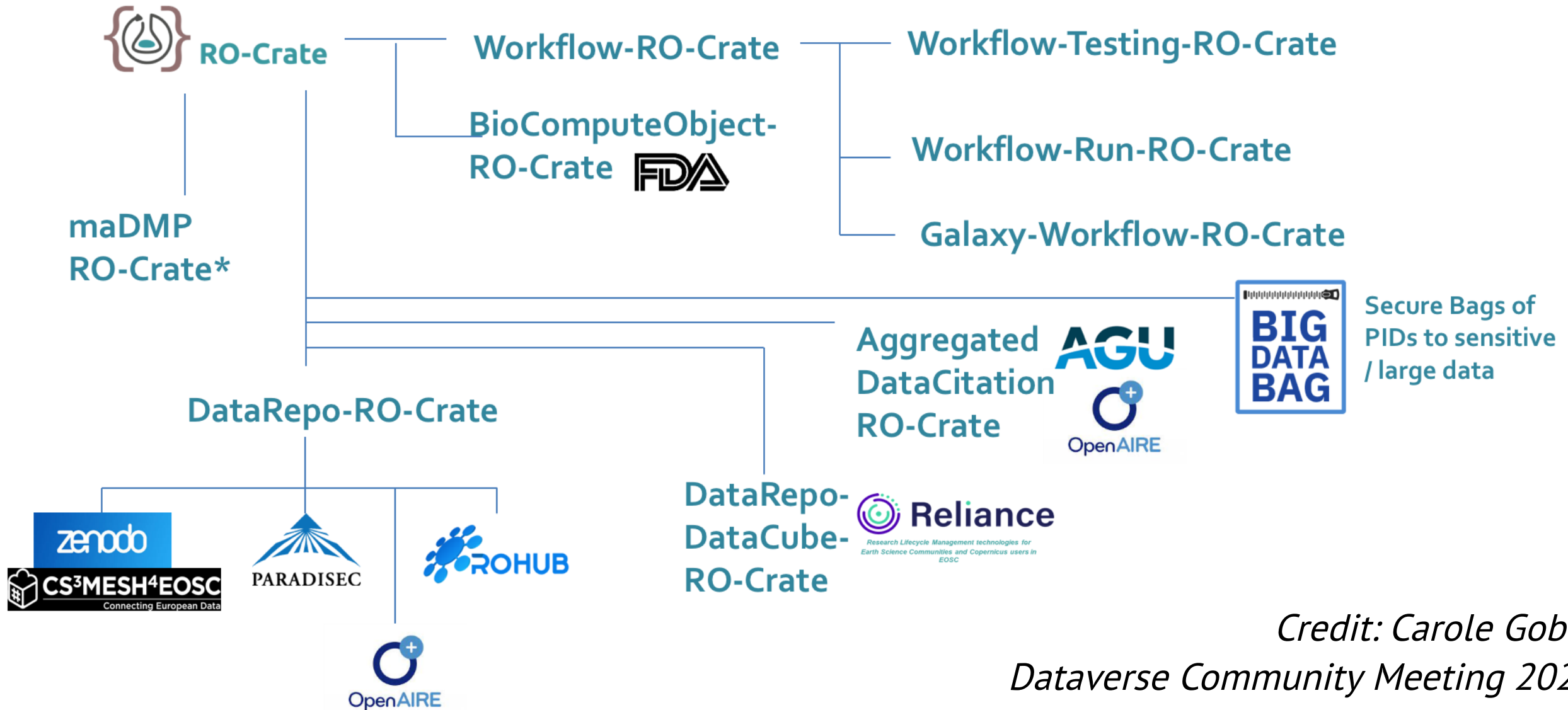
Describing workflows with RO-Crate



<https://about.workflowhub.eu/Workflow-RO-Crate/>

<https://pypi.org/project/galaxy2cwl/>

@soilandreyes



RO-Crate Community

- Peter Sefton <https://orcid.org/0000-0002-3545-944X> (co-chair)
- Stian Soiland-Reyes <https://orcid.org/0000-0001-9842-9718> (co-chair)
- Eoghan Ó Carragáin <https://orcid.org/0000-0001-8131-2150> (emeritus chair)
- Oscar Corcho <https://orcid.org/0000-0002-9260-0753>
- Daniel Garijo <https://orcid.org/0000-0003-0454-7145>
- Raul Palma <https://orcid.org/0000-0003-4289-4922>
- Frederik Coppens <https://orcid.org/0000-0001-6565-5145>
- 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 Gomez-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 Bacall <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-Navas <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 Garcia 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/0000-0002-9648-6484>
- Paul Brack <https://orcid.org/0000-0002-5432-2748>
- Milan Ojsteršek <https://orcid.org/0000-0003-1743-8300>
- Bert Driesbeke <https://orcid.org/0000-0003-0522-5674>
- Chenxu Niu <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://github.com/cedricdcc>
- Andreas Czerniak <https://orcid.org/0000-0003-3883-4169>

The RO-Crate Community is open for anyone to [join us!](#)

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

@soilandreyes