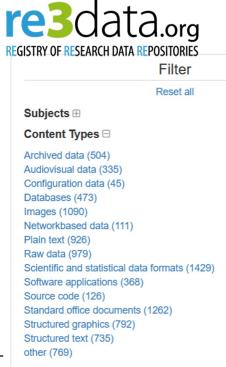
# Packaging data with detailed metadata using RO-Crate in FAIR open repositories

Peter Sefton<sup>1</sup>, Stian Soiland-Reyes<sup>2</sup> 1: The University of Queensland, Australia; 2: The University of Manchester, UK

# Is it FAIR to use all these repositories?



### https://www.re3data.org/



https://faircookbook.elixir-europe.org/

**FAIR**COOKBOOK

Findability

Unique, persistent identifier

search engine optimizatio

LEARN MORE

Infrastructure

-

LEARN MORE

## Aims of FAIR Research Objects

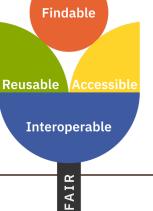
**Describe** and **package** data collections, datasets, software etc. with their **metadata** 

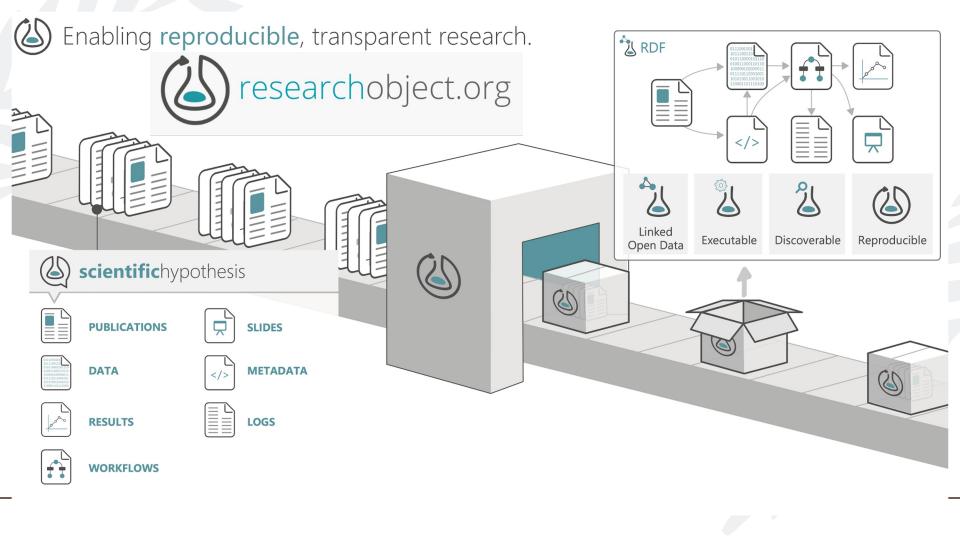
Platform-independent object exchange between repositories and

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

**RO-Crate on GitHub** 

RO-Crate 1.1	~ F	Research Object Crate (RO-Crate)
Background	Р	ermalink: https://w3id.org/ro/crate
Community		
Examples	T/	ABLE OF CONTENTS
Examples	1	What is RO-Crate?
Outreach and Publications	2	Where did RO-Crate come from?
Profiles	3	Who is it for?
RO-Crate In Use	<b>~</b> 4	How can I get started?
Specification	5	What can I use RO-Crate with?
Tools	6	RO-Crate in use
Tutorials	7	Contribute to RO-Crate community
	8	Cite RO-Crate
		a Cite RO-Crate as project/approach
		b Cite RO-Crate Specification (any version)
		c Other citations

Q

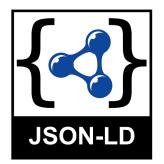
News: RO-Crate Metadata specification 1.1 released

#### What is RO-Crate?

RO-Crate is a community effort to establish a lightweight approach to packaging research data with their metadata. It is based on schema.org annotations in JSON-LD, and aims to make best-practice in formal metadata description accessible and practical for use in a wider variety of situations, from an individual researcher working with a folder of data, to large data-intensive computational research environments.

## Using common formats and vocabularies

.. extending only when needed





Pulse
Contributors
Community
Community Standards
Traffic
Commits
Code frequency
Dependency graph
Network

Forks

People

### April 30, 2023 - May 30, 2023

Active pull requests		11 Active issues		
<b>№ 13</b>	<b>្លា 1</b>	⊘ 8	⊙ <b>3</b>	
Merged pull requests	Open pull request	Closed issues	New issues	

\_ ⊱ **13** Pull requests merged by **6** people

0 -

🗿 💌 🐩 👰 💿 🔝

#### ✤ update of changelog in progress

#261 merged last week

867 deletions.

- contentUrl for direct download, add section on Converting to Attached RO-Crate #259 merged last week
- List required properties for entities #260 merged last week
- Specify how cite-as and Signposting should be used #255 merged 2 weeks ago
- . . . . . . . . . .



### **Research Object Crate community calls**

This is the rolling minutes from the <u>Research Object Crate</u> (*RO-Crate*) telcons. Feel free to join and suggest changes to the agenda or to help scribe the minutes of the call!

Regular calls are scheduled on **4th Thursdays** of the month **20:00** UTC. <u>EOSC</u>-themed calls are also scheduled on **2nd Thursdays** of the month **08:00 UTC.** (subject to change, see <u>Upcoming meetings</u> below)

Note: We follow UTC and do not observe daylight savings, so <u>double-check</u> when your country changes their clocks!

#### Upcoming meetings

- 2023-06-08 Cancelled (ELIXIR All Hands)
- 2023-06-22 20:00 UTC
- 2023-07-13 08:00 UTC (EOSC-themed)
- 2023-07-27 20:00 UTC
- 2023-08-10 Cancelled
- 2023-08-24 Cancelled
- 2023-09-14 08:00 UTC (EOSC-themed)
- 2023-09-28 20:00 UTC
- 2023-10-12 08:00 UTC (EOSC-themed) Note daylight saving changes
- 2023-10-26 20:00 UTC Note daylight saving changes
- 2023-11-09 08:00 UTC (EOSC-themed) Note daylight saving changes
- 2023-11-23 20:00 UTC Note daylight saving changes
- 2023-12-14 08:00 UTC (EOSC-themed)
- 2023-12-28 Cancelled

Tip: Import as Calendar event (ICS) (via Zoom)



### ResearchObject.org (RO)

Research Object (12)

u All versions	Found 96 results.         <
Access Right	asc. ~
□ Open (96)	June 1, 2023 (v1)         Poster         Open Access           Making workflow provenance FAIR across workflow systems with Workflow Run RO-Crate         View
File Type	🌝 Simone Leo; 💿 Laura Rodríguez-Navas; 💿 José M. Fernández; 💿 Paul De Geest; 💿 Luca Pireddu; 💿 Michael R. Crusoe; 💿 Daniel Garijo; 🕼 Iacopo Colonnelli; 🕼 Raül Sirvent; 🕼 Stian Soiland-Reyes;
□ Pdf (83)	Workflow Run RO-Crate (https://w3id.org/ro/wfrun/), is a set of profiles of RO-Crate (https://doi.org/10.3233/DS-210053) that capture workflow provenance in a lightweight FAIR data package, in order to support traceability, reproducibility and interoperable description of diverse computational
□ Zip (19)	diverse computational Uploaded on June 2, 2023
□ Pptx (16)	upioaded on June 2, 2023
🗆 Html (12)	May 30, 2023 (v2) Poster Open Access
□ Sha256 (3)	Sharing data as machine-actionable objects using RO-Crate, Bioschemas and Signposting
□ Json (2)	Stian Soiland-Reves;      Simone Leo;      Levia Jael Castro;      Peter Sefton;      Carole Goble;
□ JsonId (2)	RO-Crate (https://doi.org/10.3233/DS-210053) is a lightweight method to package research outputs along with their metadata, based on
□ Md (2)	existing Linked Data standards. Bioschemas (https://bioschemas.org/) provides metadata schemas to add structured metadata to webpages on Life Science, based on schema.
□ Mp4 (2)	Uploaded on May 31, 2023
□ Webm (2)	1 more version(s) exist for this record
Keywords	January 23, 2023 (v3) Presentation Open Access View
iteywords	

#### Sharing research artefacts as FAIR Digital Objects using RO-Crate

Carole Goble; Stian Soiland-Reyes;



#### A registry for describing, sharing and publishing scientific computational workflows

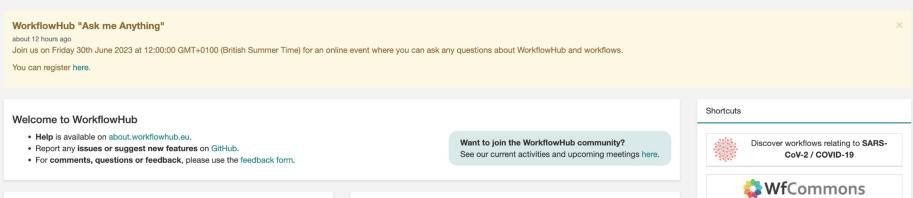
WorkflowHub aims to **facilitate discovery and re-use** of workflows in an accessible and interoperable way. This is achieved through extensive use of **open standards** and tools, including CWL, RO-Crate, Bioschemas and GA4GH's TRS API, in accordance with the **FAIR principles**.

Register

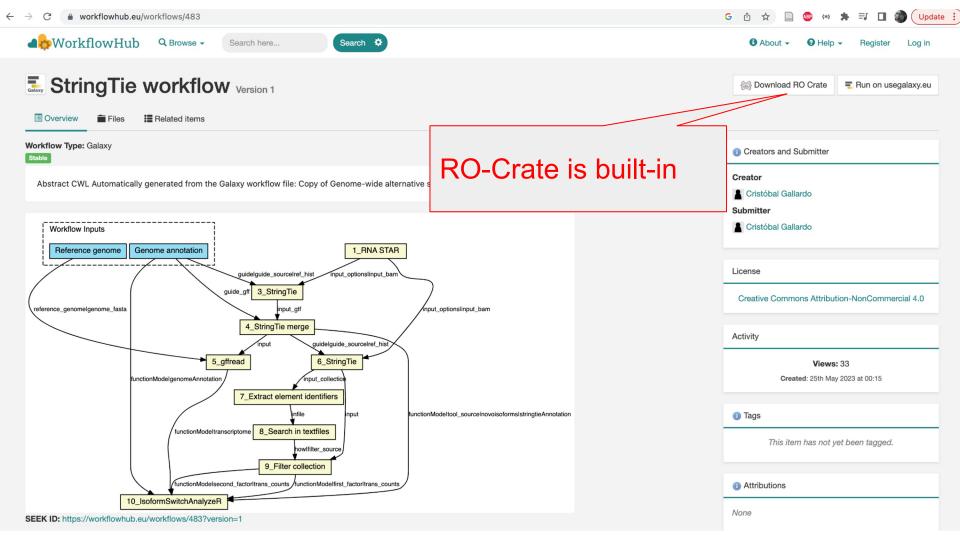
Looking for WfCommons? Click here

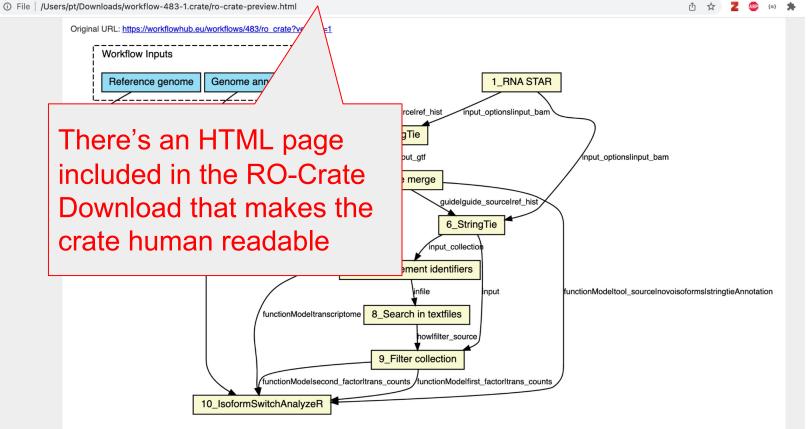
WorkflowHub supports workflows of any type in its native repository.

#### Learn more



Find content



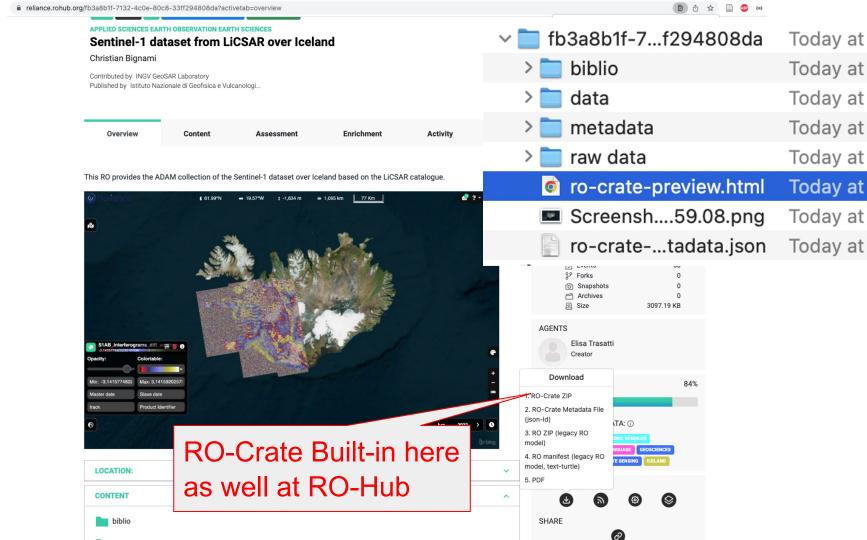


Abstract CWL Automatically generated from the Galaxy workflow file: Copy of Genome-wide alternative splicing analysis

Author Cristóbal Gallardo License CC-BY-NC-4.0

#### Contents

Main Workflow: <u>StringTie workflow</u>





#### Home /

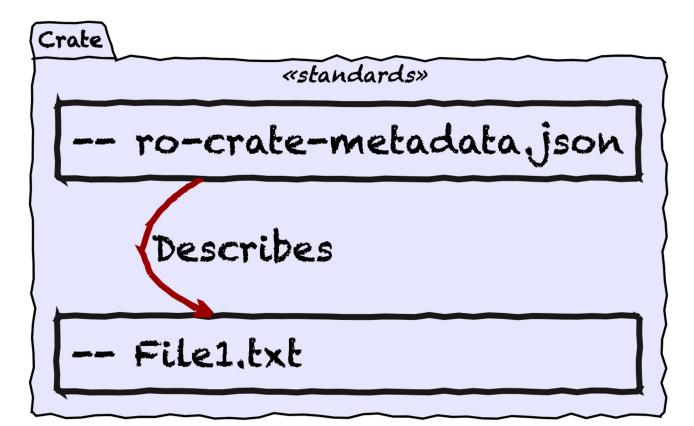
## Support offer #2: Enabling FAIR Signposting and RO-Crate for content/metadata discovery and consumption

#### Context

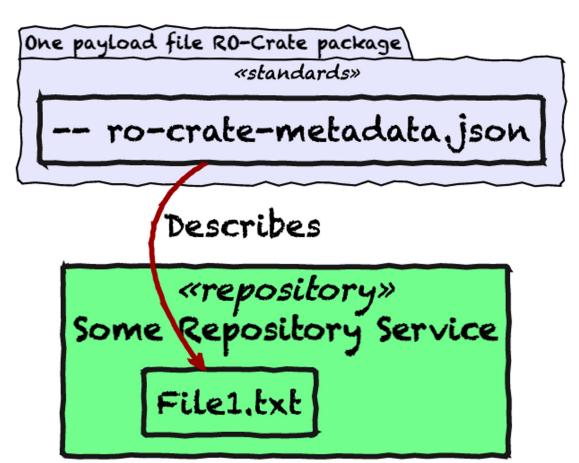
The findability of a wide range of research objects and their related metadata are central to the FAIR principles. This support action combines two successful approaches (FAIR Signposting and RO-Crate) to help ensure that research objects can be packaged up with structured metadata to support reuse and that these packages can be exposed for improved findability.

FAIR Signposting is a method to expose machine-actionable navigation links that indicate downloadable resources, types and attribution – particularly for scholarly and institutional repositories which use persistent identifiers like DOIs. Signposting makes explicit the links between a typical HTML landing page and the downloadable resources that are available for the research object described by that landing pages, including content resources and machine-readable metadata such as in RDF, although the method is technology-agnostic in terms of metadata formats. It also links to persistent identifiers, both for the research object and its authors. Signposting uses existing standards to achieve this: Web Links (RFC8288) conveyed using a simple HTTP header, HTML <link> elements, and/or Linksets (RFC9264). All link relations used in Signposting are registered in the IANA Link Relations Registry. Signposting client libraries have been

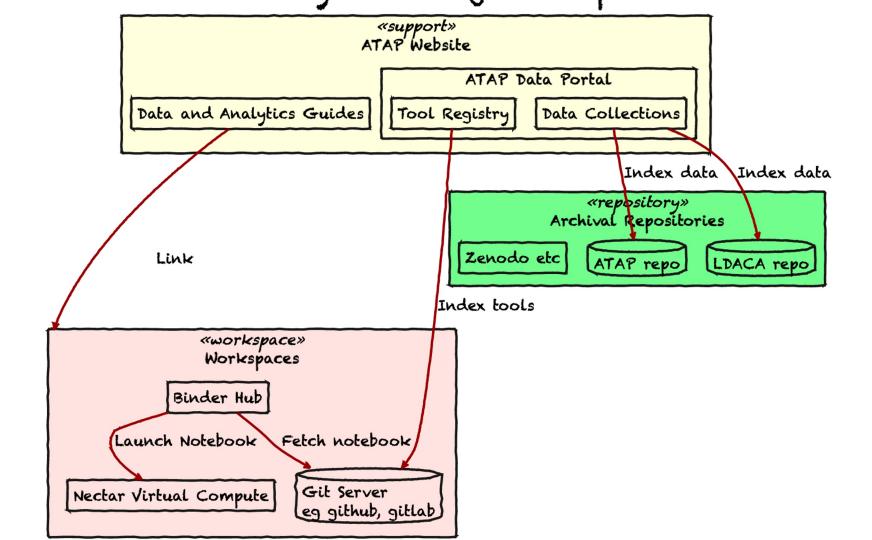
# Attached Crate



## Detached Crate



```
(i) localhost:8080/object/meta?id=arcp://name,plays/object/WoundsofCivilWar_1594
\leftarrow
   \rightarrow
      C
                                                                                                                 Ð
22
          ζ,
23
   W
             "@id": "arcp://name,plays/object/WoundsofCivilWar_1594",
24
25
             "@type": [
   W
26
               "Dataset",
               "RepositoryObject"
27
28
             ],
29
             "hasPart": {
   W
30
               "@id": "http://localhost:8080/stream?
      id=arcp://name.plays/object/WoundsofCivilWar_1594&path=Texts/WoundsofCivilWar_1594.xml"
31
             "hasMember": [
32
   W
33
34
             ٦,
35
             "conformsTo": {
               "@id": <u>"https://purl.archive.org/language-data-commons/profile#Object"</u>
36
37
             },
38
             "memberOf": {
                                                                                              References file streams
   W
39
               "@id": "arcp://name,plays"
                                                                                              from the API
40
             },
41
             "name": "Wounds of Civil War",
             "W&R_ID": "802",
42
43
             "filename": "WoundsofCivilWar_1594",
             "author": "\"Lodge Thomas\"",
44
45
             "Company of First Production": "Admiral's (Nottingham's) Men",
46
             "Date": "1588".
47
             "earliest": "1587",
48
             "latest": "1588",
             "span": "2",
49
```



### Tools

While we're mostly focusing on the <u>RO-Crate specification</u> some tools already exist for working v RO-Crates:

- Describo interactive desktop application to create, update and export RO-Crates for different profiles. (~ RC)
- Describo Online Web-based application to create RO-Crates using cloud storage (~ alpha)
- ro-crate-excel Command-line tool to help create RO-Crates from spreadsheets (~ beta)
- <u>ro-crate-html</u> HTML rendering of RO-Crate (~ beta)
- ro-crate-preview GitHub Action for ro-crate-html, e.g. publishing crates on GitHub Pages (alpha)
- ro-crate-js JavaScript/NodeJS library for RO-Crate rendering as HTML. (~ beta)
- ro-crate-ruby Ruby library to consume/produce RO-Crates (~ beta)
- ro-crate-py Python library to consume/produce RO-Crates (~ beta)
- <u>CheckMyCrate</u> Validation according to Workflow RO-Crate profile (~ *alpha*)
- galaxy2cwl Wraps Galaxy workflow as Workflow RO-Crate (~ alpha)
- ya2ro Generate RO-Crate and HTML page from YAML template with look-up of DOI/ORCID/GitHub metadata (~ prototype)
- arc-to-roc Generate RO-Crate from an Annotated Research Context (ARC), see DataPlant
- <u>ROCrate\_enrichment\_service</u> API-based metadata enrichment service for RO-Crates (~ prototype)
- rocrate-to-html Github Action to publish rocrate objects as Github Pages (~ *alpha*) (see also new\_rocrate\_to\_pages)
- FAIR-Research-Object evaluate FAIRness of Research Objects through an API (~ prototype)
- repo2crate Generate a Workflow Testing RO-Crate from a "best-practices" workflow repository
- ro-crate-java Java API for creating and modifying RO-Crate using builder pattern
- ro-crate-benchmarks Benchmarks for performance testing RO-Crate libraries
- tonkaz Tool to verify workflow reproducibility, compares RO-Crates of workflow execution results.

- signposting Python library & CLI tool for resolving PIDs as FAIR Signposting, e.g. DOIs from WorkflowHub to RO-Crate (~ beta)
- RO-Crates-and-Excel generate RO-Crate from Excel file, following the RO-Crates-and-Excel profile.
- rocrate-lang-py RO Crate Python library to help you load language data from ro-crates (~ prototype)
- <u>ROcrate-interface</u> Initial development in creating an interface between workflow languages and a LivePublication RO-crate specification (~ *prototype*)
- aiida-rocrate AiiDA plugin that allows exporting (parts) of the provenance graph as Research Object Crates. (~ *planning*)
- RO-Crate-Registry a Web-based registry of RO-Crates (assumes ZIP on http/https) (~ prototype)
- ro-crate-validator-py a modular RO-Crate validator (~ alpha)

See also applications using RO-Crate.

#### https://www.researchobject.org/ro-crate/tools/



#### A COrpus of Oz Early English (COOEE)

Name	A COrpus of Oz Early English (COOEE)	Access
Description	Material to be included had to meet with a regional and a temporal criterion. The latter required texts to have been produced between 1788 and 1900 in order to become eligible for COOEE. It was mandatory for a text to have been written in Australia, New Zealand or Norfolk Island. But in a few cases, other localities were allowed. For example, if a person who was a native Australian or who had lived in Australia for a considerable time, wrote a shipboard diary or travelled in other countries. Contains: Letters, published materials in book form, historical texts	Attribution 4.0 International (CC BY 4.0) Public Metadata  Indexed
Date Published	Not Defined	Quartered
@id @	arcp://name,cooee-corpus/corpus/root 0	Content
Author @	Clemens W. A. Fritz	Language English: 4071
Citation @	From English in Australia to Australian English	Linguistic Genre Private Written: 610
Temporal Coverage 🛛	1788-1900	Public Written: 405
Conforms To @	https://purl.archive.org/language-data-commons/profile#Collection	Government English: 195 Speech Based: 147
Identifier @	АТАР	Modality WrittenLanguage: 4071
Objects in Collection: 1357		File Formats text/plain: 2714

Text 1-001 1788 Phillip, Arthur Text 1-002 1788 Phillip, Arthur Text 1-003 1788 Phillip, Arthur Text 1-004 1788 Phillip, Arthur Text 1-005 1788 Phillip, Arthur Text 1-007 1788 Phillip, Arthur Text 1-007 1788 Phillip, Arthur Text 1-009 1788 Phillip, Arthur Text 1-009 1788 Bench of Magistrates Text 1-010 1788 Fowel, Newton

#### **Retrieve Metadata**

Download metadata Open metadata in a new window

#### Notebooks

cooee notebook

load more...



#### cooee notebook

#### **Description** A sample notebook for the cooee data

@id ø	cooee.ipynb			
Author	Foley, Ben			
Conforms To	https://purl.archive.org/language-data-commons/profile#Notebook			
Encoding Format @	application/x-ipynb+json			
Input	A COrpus of Oz Early English (COOEE)			

#### Access

Git Repository cooee Notebook Location https://github.com/Australian-Text-Analytics-Platform/cooee/blob/main/cooee.jpynb

 launch binder

#### **Notebook Viewer**

%capture import sys !{sys.executable} -m pip install -r requirements.txt

# Specify location where collection is LDACA\_API = 'https://data.atap.edu.au/api' COLLECTION\_ID = 'arcp://name,cooee-corpus/corpus/root'  $\equiv$  README.md

#### **Creating an RO-Crate**

In its simplest form, an RO-Crate is a directory tree with an ro-crate-metadata.json file at the top level that contains metadata about the other files and directories, represented by data entities. These metadata consist both of properties of the data entities themselves and of other, non-digital entities called contextual entities (representing, e.g., a person or an organization).

Suppose Alice and Bob worked on a research task together, which resulted in a manuscript written by both; additionally, Alice prepared a spreadsheet containing the experimental data, which Bob used to generate a diagram. Let's make an RO-Crate to package all this:

```
from rocrate.rocrate import ROCrate

crate = ROCrate()
paper = crate.add_file("exp/paper.pdf", properties={
    "name": "manuscript",
    "encodingFormat": "application/pdf"
})
table = crate.add_file("exp/results.csv", properties={
    "name": "experimental data",
    "encodingFormat": "text/csv"
})
```

diagram = crate.add file("exp/diagram.svg", dest path="images/figure.svg", prop

Ø

Q

#### $\equiv$ README.md

#### **Crate initialization**

The rocrate init command explores a directory tree and generates an RO-Crate metadata file (ro-crate-metadata.json) listing all files and directories as File and Dataset entities, respectively.

```
$ rocrate init --help
Usage: rocrate init [OPTIONS]
```

```
Options:

--gen-preview

-e, --exclude CSV

-c, --crate-dir PATH

--help Show this message and exit.
```

The command acts on the current directory, unless the -c option is specified. The metadata file is added (overwritten if present) to the directory at the top level, turning it into an RO-Crate.

#### Adding items to the crate

The rocrate add command allows to add workflows and other entity types (currently

0

#### E README.md

#### Usage

Import the R0Crate class and create a new empty crate with default configurations:

const {ROCrate} = require('ro-crate'); const crate = new ROCrate();

The ROCrate constructor accepts two optional arguments:

const fs = require('fs');

```
// load existing metadata
const data = JSON.parse(fs.readFileSync('ro-crate-metadata.json', 'utf8'));
```

// create a crate using the existing data and // configure the crate to return a property of an Entity as an array and resolve linked entity as ne const crate = new ROCrate(data, { array: true, link: true });

0

#### To add an Entity to the crate:

#### // A license

const license = {
 '@id': 'https://creativecommons.org/licenses/by/4.0/',
 '@type': 'CreativeWork',
 'description': 'Attribution 4.0 International (CC BY 4.0) ...',
 'name': 'CC BY 4.0'

};

// add the license as an unconnected Entity
crate.addEntity(license);

// add the license to the root dataset crate.rootDataset.license = {'@id': license['@id']}; // or alternatively, add a new entity directly into a property of other entity : crate.rootDataset.license = license;

#### Use an entity just like a normal object:

let lic = create.getEntity(license['@id']); console.log(lic.name); // prints 'CC BY 4.0'; // set a property directly lic.name = 'CC BY 4.0 dummy'; // or with the setProperty method crate.setProperty(license['@id'], 'name', 'CC BY 4.0 dummy');

console.log(lic.name); // prints 'CC BY 4.0 dummy';

	А	В	С	D	E	F	G
1	@id	@type	name	description	contentSize	dateModified	encodingFormat
2	LICENSE	File	LICENSE		35149	2023-06-01T13:52:12+10	{"@id":"https://www.nati
3	README.md	File	README.md		188	2023-06-01T13:52:12+10	{"@id":"https://www.nati
4	additional-ro-crate-metad	File	additional-ro-crate-metad		95008	2023-06-09T16:05:03+10	{"@id":"https://www.nati
5	metadata_279_plays_and	File	metadata_279_plays_and		53377	2023-06-01T13:53:21+10	{"@id":"https://www.nati
6	Texts/1EdwardIV_1599.xr	File	1EdwardIV_1599.xml		532860	2023-06-01T10:19:28+10	{"@id":"https://www.nati
7	Texts/1FairMaidoftheWes	File	1FairMaidoftheWest_163		392256	2023-06-01T10:19:28+10	{"@id":"https://www.nati
8	Texts/1HenryIV_1598.xml	File	1HenryIV_1598.xml		433218	2023-06-01T10:19:28+10	{"@id":"https://www.nati
9	Texts/1HenryIV_1623.xml	File	1HenryIV_1623.xml		435551	2023-06-01T10:19:28+10	{"@id":"https://www.nati
.0	Texts/1HenryVI_1623.xml	File	1HenryVI_1623.xml		414846	2023-06-01T10:19:28+10	:00
.1	Texts/1lfyouKnowNotMe_	File	1lfyouKnowNotMe_1605		261385	2023-06-01T10:19:28+10	{"@id":"https://www.nati
2	Texts/1Jeronimo_1605.xr	File	1Jeronimo_1605.xml		183793	2023-06-01T10:19:28+10	:00
.3	Texts/1Selimus_SilentReg	File	1Selimus_SilentReg_1638		312327	2023-06-01T10:19:28+10	:00
.4	Texts/1SirJohnOldcastle_1	File	1SirJohnOldcastle_1600.>		476514	2023-06-01T10:19:28+10	{"@id":"https://www.nati
.5	Texts/1Tamburlaine_Silen	File	1Tamburlaine_SilentReg_		320252	2023-06-01T10:19:28+10	{"@id":"https://www.nati
.6	Texts/1TwoAngryWomenc	File	1TwoAngryWomenofAbin		505066	2023-06-01T10:19:28+10	{"@id":"https://www.nati
.7	Texts/2 HenryIV_1623.xm	File	2 HenryIV_1623.xml		455611	2023-06-01T10:19:28+10	{"@id":"https://www.nati
.8	Texts/2EdwardIV_1599.xr	File	2EdwardIV_1599.xml		554468	2023-06-01T10:19:28+10	{"@id":"https://www.nati
.9	Texts/2FairMaidoftheWes	File	2FairMaidoftheWest_163		479066	2023-06-01T10:19:28+10	{"@id":"https://www.nati
0	Texts/2HenryIV_1600_QB	File	2HenryIV_1600_QB.xml		325006	2023-06-01T10:19:28+10	{"@id":"https://www.nati
1	Texts/2HenryVI_1594.xml	File	2HenryVI_1594.xml		364444	2023-06-01T10:19:29+10	{"@id":"https://www.nati
2	Texts/2HenryVI_1623.xml	File	2HenryVI_1623.xml		478881	2023-06-01T10:19:29+10	:00
3	Texts/2HonestWhore_163	File	2HonestWhore_1630.xml		474458	2023-06-01T10:19:29+10	{"@id":"https://www.nati
4	Texts/2IfYouKnowNotMe_	File	2IfYouKnowNotMe_1606		528143	2023-06-01T10:19:29+10	{"@id":"https://www.nati
5	Texts/2Tamburlaine_1590	File	2Tamburlaine_1590.xml		321351	2023-06-01T10:19:29+10	:00
6	Texts/3HenryVI_1595.xml	File	3HenryVI_1595.xml		338051	2023-06-01T10:19:29+10	:00
7	Texts/3HenryVI_1623.xml	File	3HenryVI_1623.xml		479113	2023-06-01T10:19:29+10	:00
8	Texts/ALarumforLondon_:	File	ALarumforLondon_1602.		294097	2023-06-01T10:19:29+10	{"@id":"https://www.nati
9	Texts/Aglaura_SilentReg_1	File	Aglaura_SilentReg_1638.x		270200	2023-06-01T10:19:29+10	{"@id":"https://www.nati

← → C	language-research-technology.github.io/crate-o/#/?id=%2523Person-1	다 🏠 🛧 📄 🐵 🖶 🖬 🌒 🗄					
Crate-O							
File 🗸	Profile: Language Data Commons top level Collection (corpus)	~					
Selected Direc	Selected Directory: corpus-tools-example-plays						
▲Root Dataset	/ Hugh Craig						
@id 🕖	https://orcid.org/0000-0002-9336-1678	Reverse Links All Entities					
@type 🕖	Person	C Dataset ./					
	+ Select						
Name 🕖	Hugh Craig						
Description ()	+ TextArea						
Affiliation ()	+ Organization						
B	N.	Obau all X					



RO-Crate 1.1 Background

Community

Examples

**Outreach and Publications** 

Profiles

**RO-Crate In Use** 

Specification

Tools

Tutorials

 $\sim$ 

 $\sim$ 

#### Team

The RO-Crate team is:

- 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)

Join us!

- 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 0
- 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

### **RO-Crate Community**

TABLE OF CONTENTS

#### 1 Team

#### 2 Meetings

#### 3 Mailing list

4 Slack Chat

5 Code of Conduct

6 Contributing

7 Open Source