Five ways RO-Crate data packages are important for repositories

Peter Sefton*, Stian Soiland-Reyes**

*University of Queensland, Australia; **The University of Manchester, UK

1. Uploading of complex multi-file objects …

… is already supported even if the repo does not speak RO-Crate

RO-Crate Metadata Document

```
|
|- ro-crate-metadata.json
|-- 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"
    },
```

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

File | /Users/pt/Do... Update

Download all the metadata for in JSON-LD format

Download: Cute puppy

@id	2021-04-08 07.58.17.jpg
name [?]	Cute puppy
@type	File

webinar-886835....ics Show all

RO-Crate Submission API

Submitting Workflows

To submit a new workflow to WorkflowHub as an RO-Crate, POST a multipart request to the `/workflows` endpoint with the following parameters:

- `ro_crate` - The RO-Crate Zip file
- `workflow[project_ids][]` - The ID of the Team that should own the RO-Crate. Can be included multiple times to be owned by more than one team.

To add a new version to an existing workflow, POST to e.g. `/workflows/123/create_version` where `123` is the workflow ID.

To change the policy and other metadata of the workflow, see the JSON API docs

Examples

The following examples assume authentication via API token. See the docs for details on other authentication methods.

They also assume your RO-Crate is in a file named `my_ro_crate.crate.zip`, and you want to add it to a Team with ID `1234`.

Curl

```
curl -X POST -H "Authorization: Token YOUR_TOKEN_HERE" \
     -F workflow[project_ids][]=1234 \
     -F ro_crate=@my_ro_crate.crate.zip https://workflowhub.eu/workflows
```

Python + requests

```
import requests
payload = { 'ro_crate': ('my_ro_crate.crate.zip ', open('my_ro_crate.crate.zip', 'rb')),
            'workflow[project_ids][]': (None, '1234') }
headers = { 'authorization': 'Token YOUR_TOKEN_HERE' }

response = requests.post('https://workflowhub.eu/workflows', files=payload, headers=headers)
```

2. RO-Crate is a packaging format suitable for downloads

2.

Gazetteer of Historical Australian Places

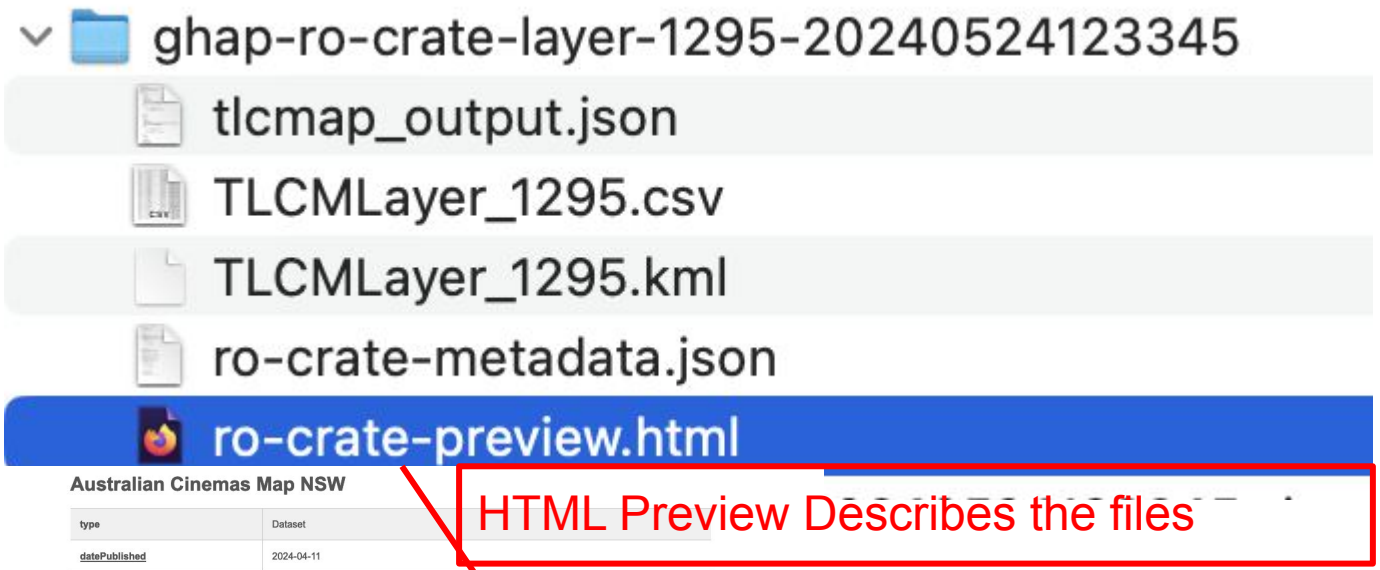
GHAP

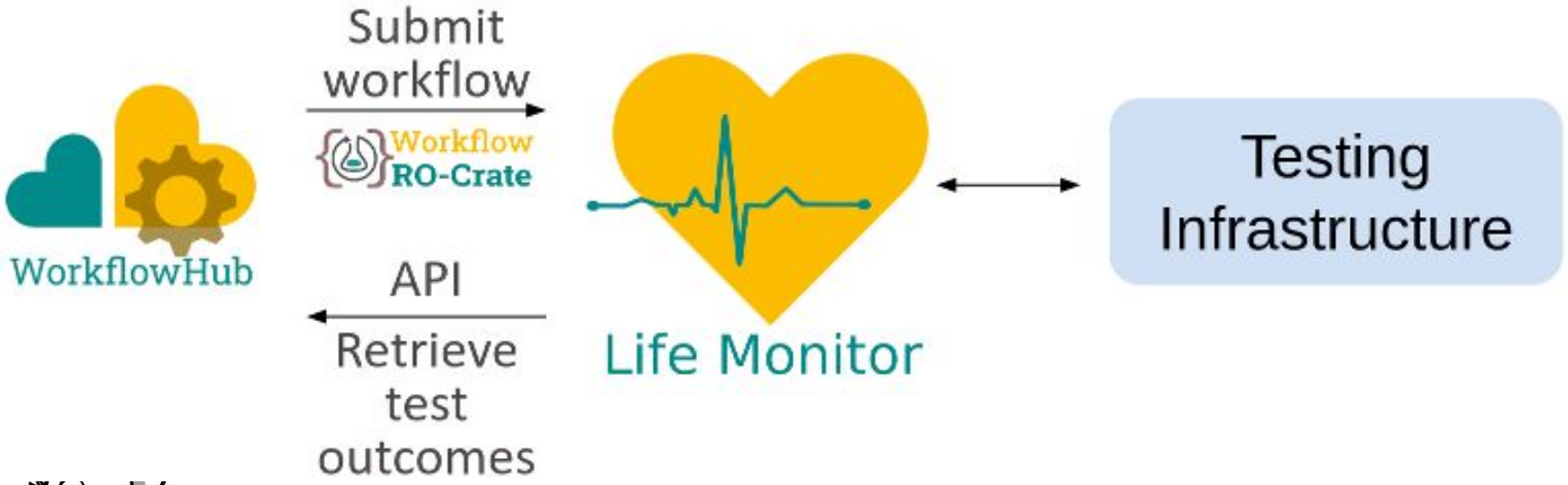
TLC MAP

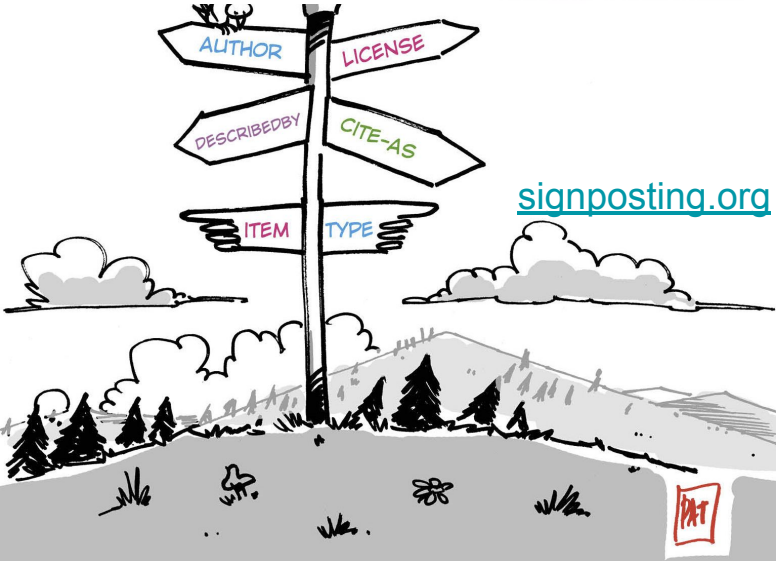
Search Layers Help

Log in

Enter search | Contains | GHAP ANPS NCG | Search | Advanced Search


ghap-ro-crate-layer-1295-20240524123345
tlcmap_output.json
TLCMLayer_1295.csv
TLCMLayer_1295.kml
ro-crate-metadata.json
ro-crate-preview.html
Australian Cinemas Map NSW
type
Dataset
datePublished
2024-04-11
name
Australian Cinemas Map NSW
description
An extract of the NSW sites from the 'Australian Cinemas Map; 1948-1971' layer, which was contributed by Peter Mason and created by Richard Maltby and Michael Walsh.
creator
Hugh Craig
publisher
Hugh Craig
url
https://ghap.tlcmap.org/publicdatasets/1295
keywords
Site
Files
CSV export of Australian Cinemas Map NSW
name
CSV export of Australian Cinemas Map NSW
description
CSV export of the layer data
encodingFormat
text/csv
File
CSV export of Australian Cinemas Map NSW
HTML Preview Describes the files

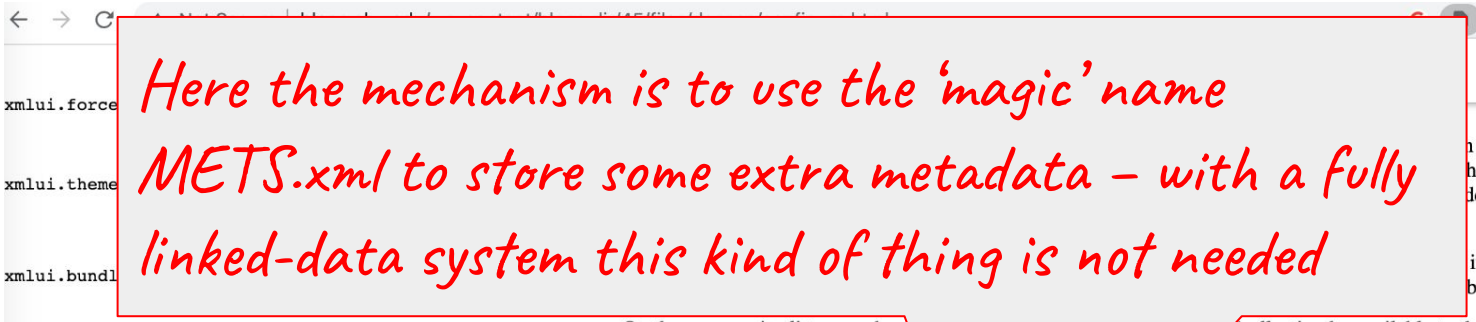




signposting.org

Enable programmatic downloads – include the metadata and its extensions

3. Less user interface customisation will be needed for different types of metadata



xmlui.force	
xmlui.theme	
xmlui.bundl	
xmlui.community-list.render.full	True
xmlui.community-list.cache	12 hours
xmlui.bitstream.mods	true
xmlui.bitstream.mets	true

. If set to true, then you need to ensure

submitting a request add the HTTP parameter
he any other configured theme. Note that this is a
development and debugging it should be turned of

item through the administrative interface. If the
be shown to the user as an option.

On the community-list page shou /collection be available to the theme. This parameter defaults to true, but if yo
are experiencing performance prob you should experiment with turning this option off.

Normally, Manakin will fully verify a cache copy. This means that when the community-list page is viewed the databas
is queried for each community/collec ta has been modified. This can be expensive for repositories with a large community
tree. To help solve this problem you ca assumed valued for a specific set of time. The downside of this is that new or editing
communities/collections may not show a period of time.

Optionally you may configure Manakin age of metadata stored as a bitstream. The MODS metadata file must be inside the
"METADATA" bundle and named either ml. If this option is set to `true` and the bitstream is present then it is made available to the theme fo
display.

Optionally you may configure Manakin to take advantage of metadata stored as a bitstream. The METS metadata file must be inside the
"METADATA" bundle and named either METS.xml. If this option is set to `true` and the bitstream is present then the stored METS file is merged wit
the METS file generated by Manakin for each item. Thus if the bitstream contains a `dmdSec` then there will be two `dmdSec` one from the bitstream and
another generated from the Dublin Core stored inside the database.

Configuring Themes and Aspects

Show more >

Implementation of a reproducible pipeline for forecasting sea ice

Vanessa Stoeckl, University of Freiburg, Germany, Anne Fouilloux, Simula Research Laboratory, Norway, Jean Iaquinta, University of Oslo, Norway, Bjoern Gruening, University of Freiburg, Germany, Alejandro Coca-Castro, The Alan Turing Institute, UK

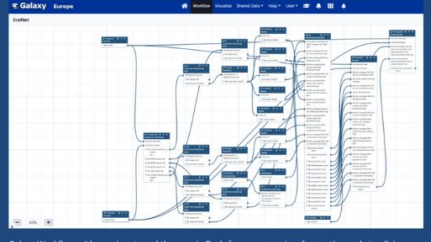
Introduction

The **Environmental Data Science book** (EDS book[1]) is a pan-european community-driven resource that leverages Pangeo software stack[2], Jupyter notebooks[3] and Research Objects[4] for FAIR environmental data science.
EDS book allows researchers to co-design collaboratively and openly review, and curate interactive, shareable and reproducible executable notebooks. Clear guidelines for writing modular and reusable Jupyter notebooks are provided but the need for understanding a specific programming language (Python, Julia, R) remains. This is clearly a barrier for interdisciplinary research and to overcome this limitation, individual and modular **Galaxy Tools** can be created out of the Jupyter notebook, for each section.
The different Galaxy Tools can then be reused by anyone from Galaxy portals to derive new innovative and fully annotated reproducible workflows.

Due to global warming, Arctic sea ice is shrinking at a rate of 12.6 % per decade [7]. This has consequences for Arctic communities and ecosystems.

Use Case

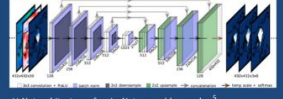
The EDS book Jupyter notebook written in Python reproduces the scientific results of the Nature Communications paper titled "Seasonal Arctic sea ice forecasting with probabilistic deep learning[5,6]". It demonstrates the Sea Ice Forecasting System by forecasting the year 2020. By creating a Galaxy Workflow from the EDS book, we are able to forecast a user defined target period. The resulting workflow uses existing Galaxy Tools and mainly three newly developed Galaxy Tools for preprocessing input data, forecasting seasonal arctic sea ice with pretrained probabilistic deep learning models and visualizing the results.
A reproducible workflow, the ability to easily replace or further expand any tool and the availability are the main benefits of implementing tools in Galaxy.



Galaxy Workflow with user inputs and three main Tools for preprocessing, forecasting and visualizing IceNet data, surrounded by Tools for downloading and preparing datasets

Sea Ice Forecasting System

The probabilistic, deep learning Sea Ice Forecasting system used in the paper[5] is called **IceNet**. It comprises an ensemble of 25 U-net networks and uses the knowledge gained from both climate simulations from 1850 to 2100 and observational data from 1979 to 2011 to predict monthly-averaged sea ice concentration maps at a high resolution of 25 kilometres for up to 6 months into the future. IceNet not only expands the horizons of accurate sea ice prediction, but also surpasses the performance of state-of-the-art dynamical models.



U-Net architecture of an IceNet ensemble member[5]

References

[1] https://eds-book.org/
[2] https://pangeo.io/
[3] http://jupyter.org/
[4] https://www.researchobject.org/
[5] https://www.nature.com/articles/s41467-021-25757-4
[6] Sea ice forecasting using IceNet — Environmental Data Science Book (eds-book.org)
[7] Arctic Sea Ice Minimum Extent | Vital Signs – Climate Change: Vital Signs of the Planet (nasa.gov)

eosc | EuroScienceGateway

CONTENT

- biblio
  - The Galaxy platform for accessible, reproducible and collabo...
- input
  - Sea ice forecasting using IceNet (Jupyter Notebook) publishe...

AGENTS

Anne Fouilloux
Creator

COMPLETENESS 0%

DISCOVERED METADATA:

SOCIOLOGY | JOB MARKET | SOFTWARE | COMPUTER SCIENCE | INTERNET | PHYSICAL GEOGRAPHY AND ENVIRONMENTAL G... | EARTH SCIENCES | CLIMATE CHANGE | WIRELESS TECHNOLOGY | SCIENCE AND TECHNOLOGY | WEATHER | HARDWARE | OIL AND GAS - UPSTREAM ACTIVITIES | OCEANOGRAPHY | GEOSCIENCES | GEOPHYSICS | GEOSCIENCES | ARCTIC ZONE

No need to invent (completely) new file formats anymore!

TheELNFileFormat / SPECIFICATION.md

Preview | Code | Blame | 187 lines (146 loc) · 6.19 KB | Raw

An up to date version of this document can be accessed at: https://github.com/TheELNConsortium/eln-file-format

This archive format is basically a zipped RO-Crate, with a `.eln` file extension.

Structure of the archive

Inside a .eln file, there MUST be a folder that will contain the rest of the data. The name of the folder SHOULD be the same as the archive name. This folder at root prevents issues when opening the file as a zip file and getting archived files extracted in the current directory, possibly overwriting other files, and probably polluting the current directory. There MUST be only one folder at the root of the archive.

Inside that root folder, there MUST be a file named `ro-crate-metadata.json`. This file follows the RO-Crate 1.1+ Specification.

The rest of the archive is composed of 0 or more folders that each describe one experiment or coherent set of data. Thus, the ELN archive can accomodate one or several experimental set of data.

4. The availability of RO-Crate editing tools opens the way for repository software to focus on access and discoverability

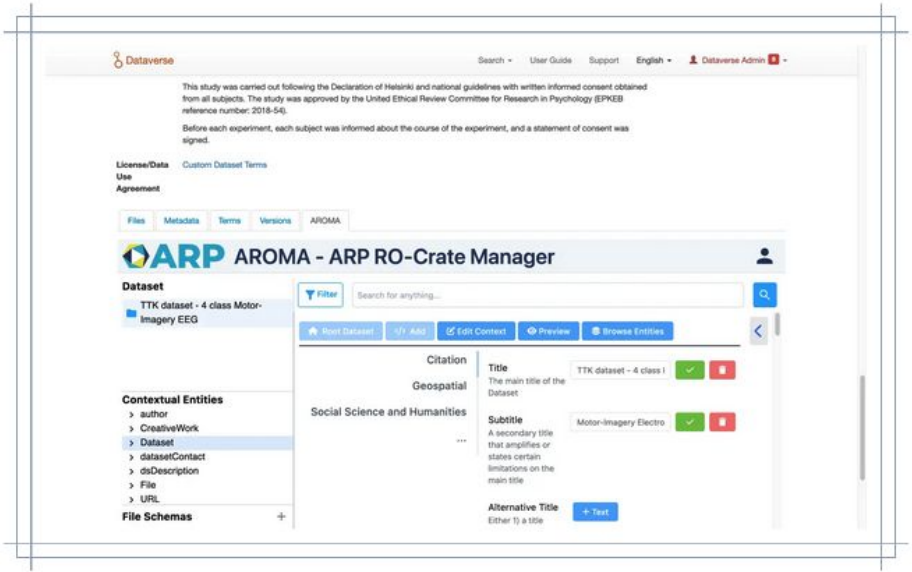
AROMA - the ARP RO-Crate Manager

AROMA - the ARP RO-Crate Manager - forms an integral part of the ELKH ARP ↗ initiative in Hungary, dedicated to delivering comprehensive data repository and metadata services to the Hungarian research community.

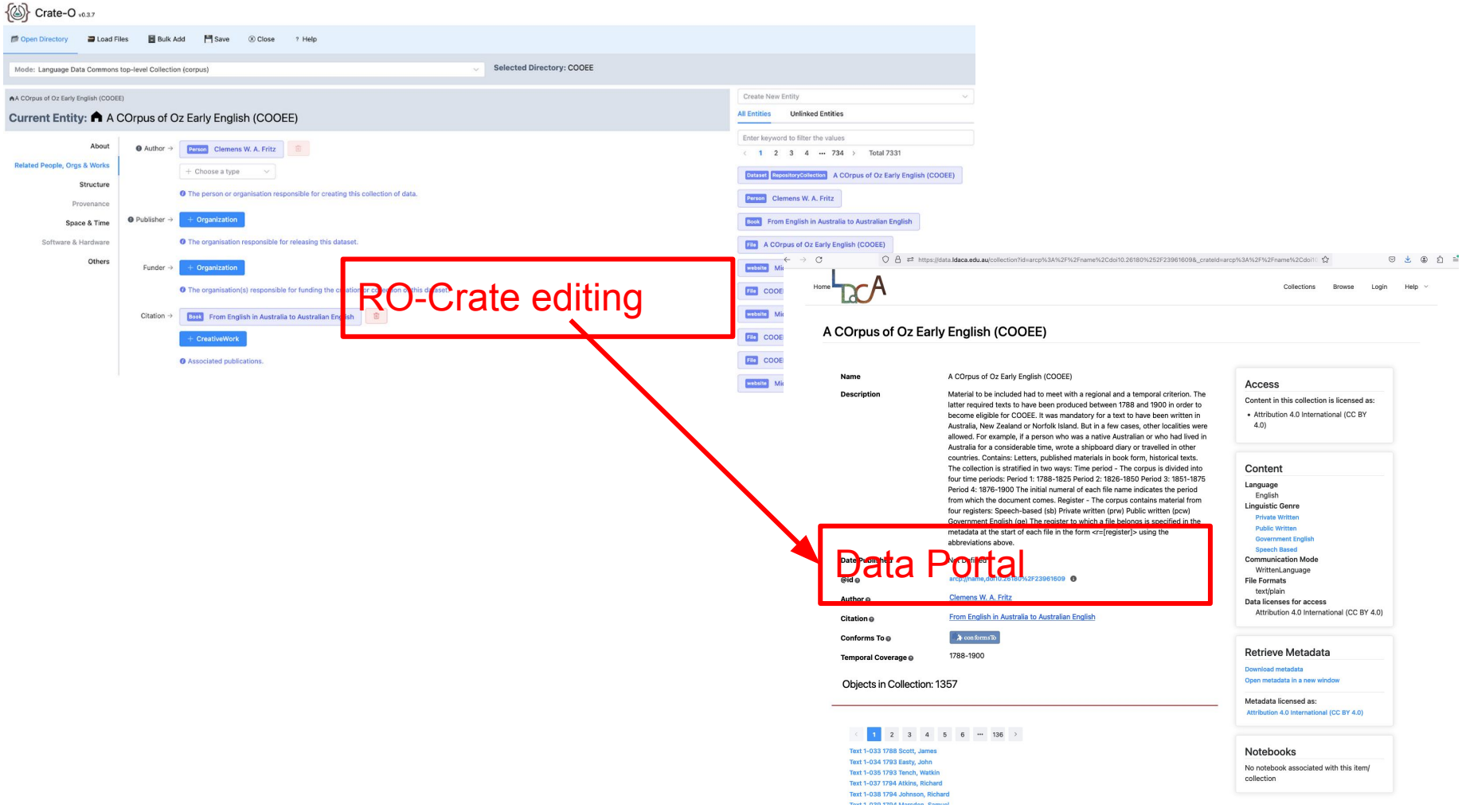
AROMA serves as a powerful extension to the Dataverse data repository system, enhancing its capabilities by enabling the description of Dataverse datasets and their internal files utilizing the RO-Crate format. AROMA ensures seamless synchronization between Dataverse and the RO-Crate metadata JSON based on schemas provided by the Dataverse metadatablocks. However, AROMA goes a step further by allowing the specification of additional metadata not only at the dataset level but also at the file level, something that isn't supported by Dataverse. To add to its flexibility and user convenience, AROMA can be employed either as an integrated component within Dataverse or as a standalone web application.

AROMA is built using the [Describo Crate Builder react component]

Aroma inside Dataverse.



2 / 2


Crate-O v0.3.7
Open Directory
Load Files
Bulk Add
Save
Close
Help
Mode: Language Data Commons top-level Collection (corpus)
Selected Directory: COOEE
A COrpus of Oz Early English (COOEE)
Current Entity: A COrpus of Oz Early English (COOEE)
About
Related People, Orgs & Works
Structure
Provenance
Space & Time
Software & Hardware
Others
Author
Person
Clemens W. A. Fritz
Choose a type
The person or organisation responsible for creating this collection of data.
Publisher
Organization
The organisation responsible for releasing this dataset.
Funder
Organization
Citation
Book
From English in Australia to Australian English
CreativeWork
Associated publications.
Create New Entity
All Entities
Unlinked Entities
Enter keyword to filter the values
Total 7331
Dataset
RepositoryCollection
A COrpus of Oz Early English (COOEE)
Person
Clemens W. A. Fritz
Book
From English in Australia to Australian English
File
A COrpus of Oz Early English (COOEE)
RO-Crate editing
Data Portal
Home
Collections
Browse
Login
Help
A COrpus of Oz Early English (COOEE)
Name
A COrpus of Oz Early English (COOEE)
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. The collection is stratified in two ways: Time period - The corpus is divided into four time periods: Period 1: 1788-1825 Period 2: 1826-1850 Period 3: 1851-1875 Period 4: 1876-1900 The initial numeral of each file name indicates the period from which the document comes. Register - The corpus contains material from four registers: Speech-based (sb) Private written (prw) Public written (pcw) Government English (ge) The register to which a file belongs is specified in the metadata at the start of each file in the form <=[register]> using the abbreviations above.
Date Published
@id
Author
Clemens W. A. Fritz
Citation
From English in Australia to Australian English
Conforms To
conformsTo
Temporal Coverage
1788-1900
Objects in Collection: 1357
Text 1-033 1788 Scott, James
Text 1-034 1793 Easty, John
Text 1-035 1793 Tench, Watkin
Text 1-037 1794 Atkins, Richard
Text 1-038 1794 Johnson, Richard
Access
Content in this collection is licensed as:
Attribution 4.0 International (CC BY 4.0)
Content
Language
English
Linguistic Genre
Private Written
Public Written
Government English
Speech Based
Communication Mode
WrittenLanguage
File Formats
text/plain
Data licenses for access
Attribution 4.0 International (CC BY 4.0)
Retrieve Metadata
Download metadata
Open metadata in a new window
Metadata licensed as:
Attribution 4.0 International (CC BY 4.0)
Notebooks
No notebook associated with this item/ collection

show fields included in search

Collection

1 2 3 4

Filter

- A COrpus of Oz Early English (COOEE) 4071
- Australian Corpus of English 3400
- Braided Channels 395
- AustLit 268
- ICE: S1A: Conversations 202

Access

Filter

- Attribution 4.0 International (CC BY 4.0) 8720
- Attribution-NoDerivs 3.0 Australia (CC BY-ND 3.0 AU) 396
- Data License for AustLit 269

Record Type 12

Language 1

Filtering by: license | Data License for AustLit × Total: 269 Index entries (Collections, Objects, Files and Notebooks)

RESTART SEARCH | Sort by: Relevance | Order by: Descending

1 2 3 4 5 6 ... 27

Convict Once

Type: RepositoryObject

Language: English

Member of: AustLit

Search Score: 1

See more

Login Required

convict-once-origin

Type: File DerivedMaterial

Member of: AustLit

XML derivative of the original

Search Score: 1

See more

Login Required

convict-once-raw

Type: File DerivedMaterial

Member of: AustLit

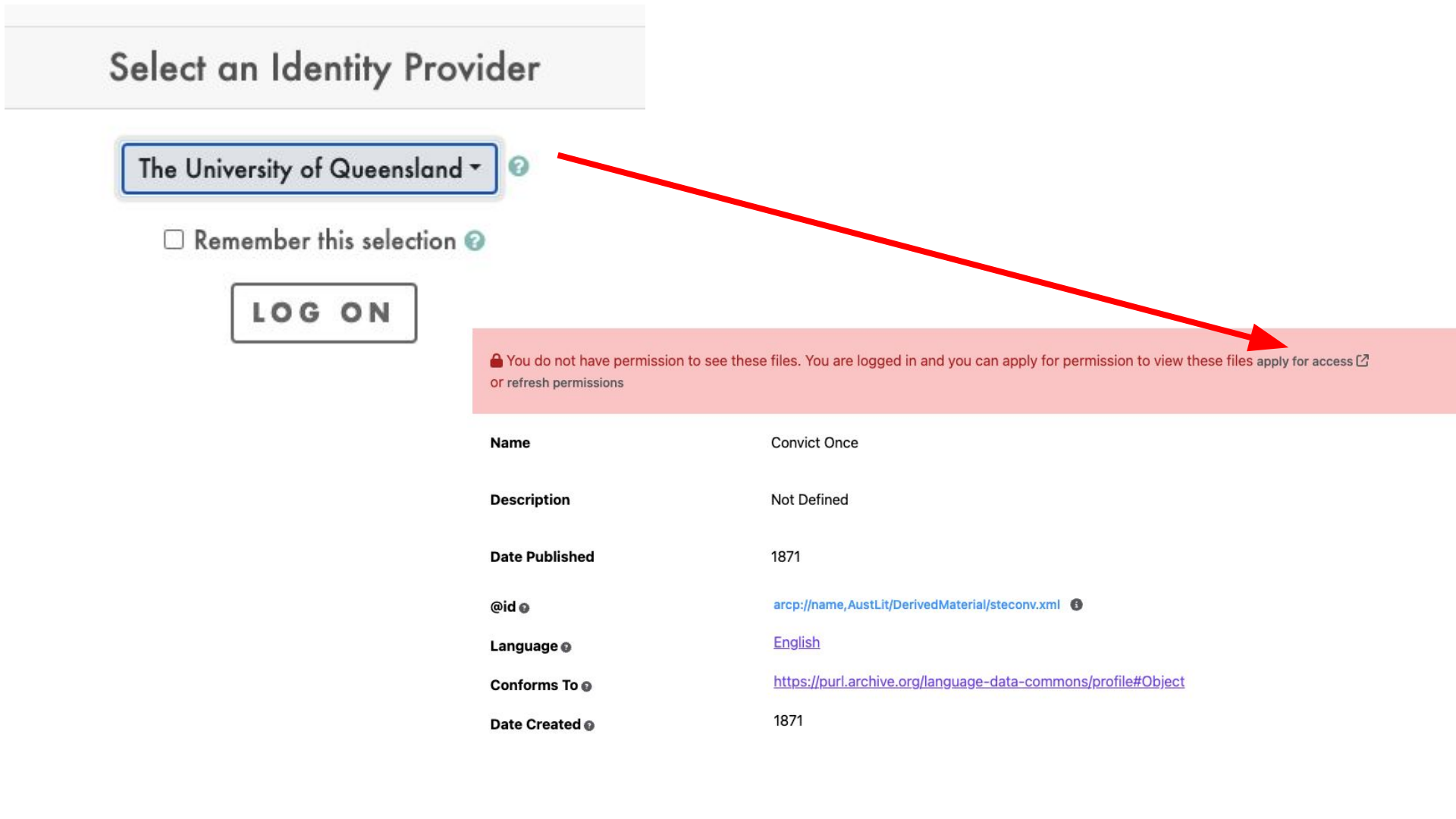
TXT derivative of the original work, contains TEI markup

Search Score: 1

See more

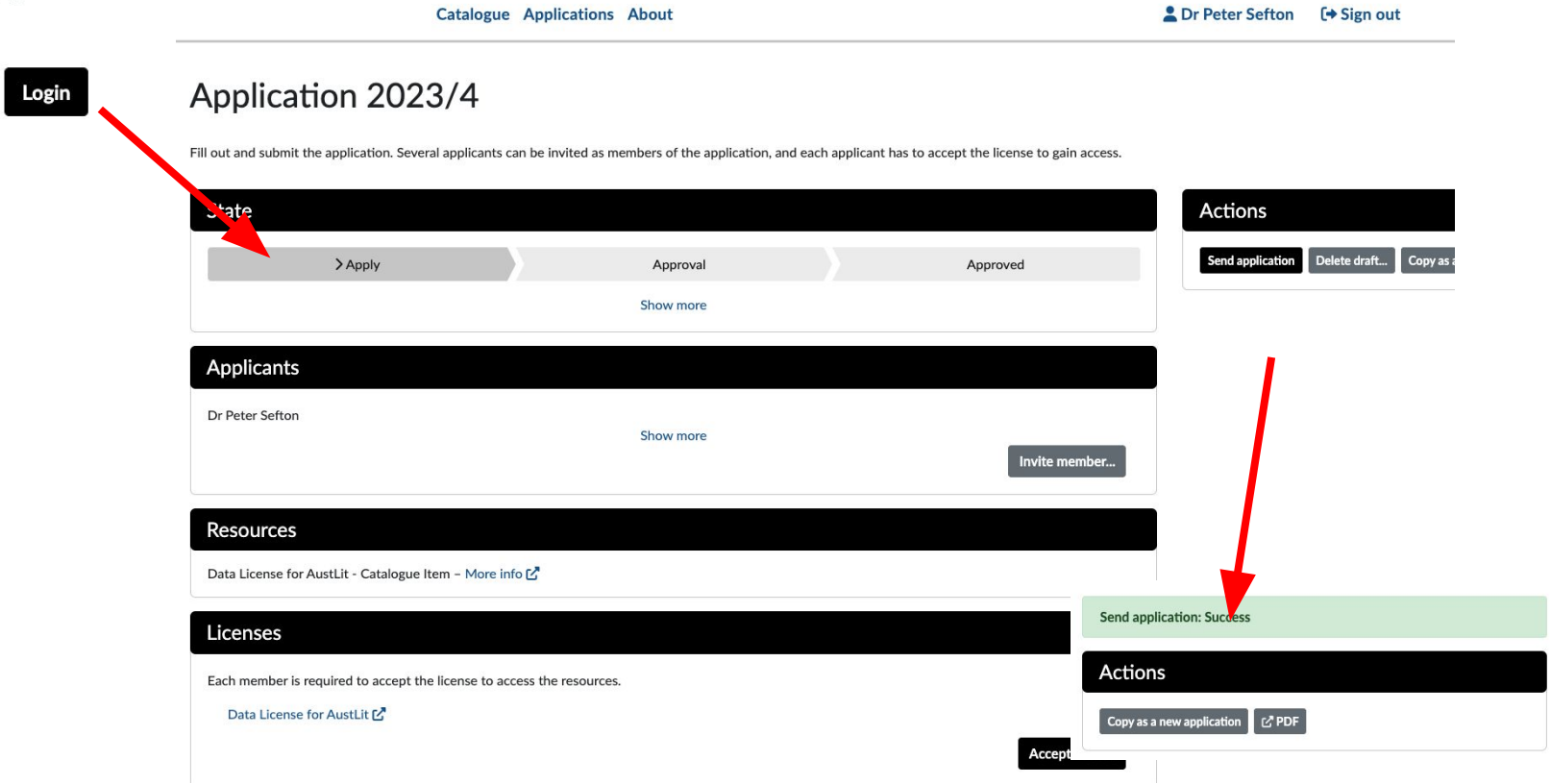
Login Required

Portal shows that accessing this data requires authorization. First step: Log in.


Select an Identity Provider
The University of Queensland
Remember this selection
LOG ON
You do not have permission to see these files. You are logged in and you can apply for permission to view these files apply for access
or refresh permissions
Name
Convict Once
Description
Not Defined
Date Published
1871
@id
arcp://name,AustLit/DerivedMaterial/steconv.xml
Language
English
Conforms To
https://purl.archive.org/language-data-commons/profile#Object
Date Created
1871

Welcome to LDaCA REMS

This is the Resource Entitlement Management System for LDaCA Program. More information is available on About page.
Please, login to access REMS.



State

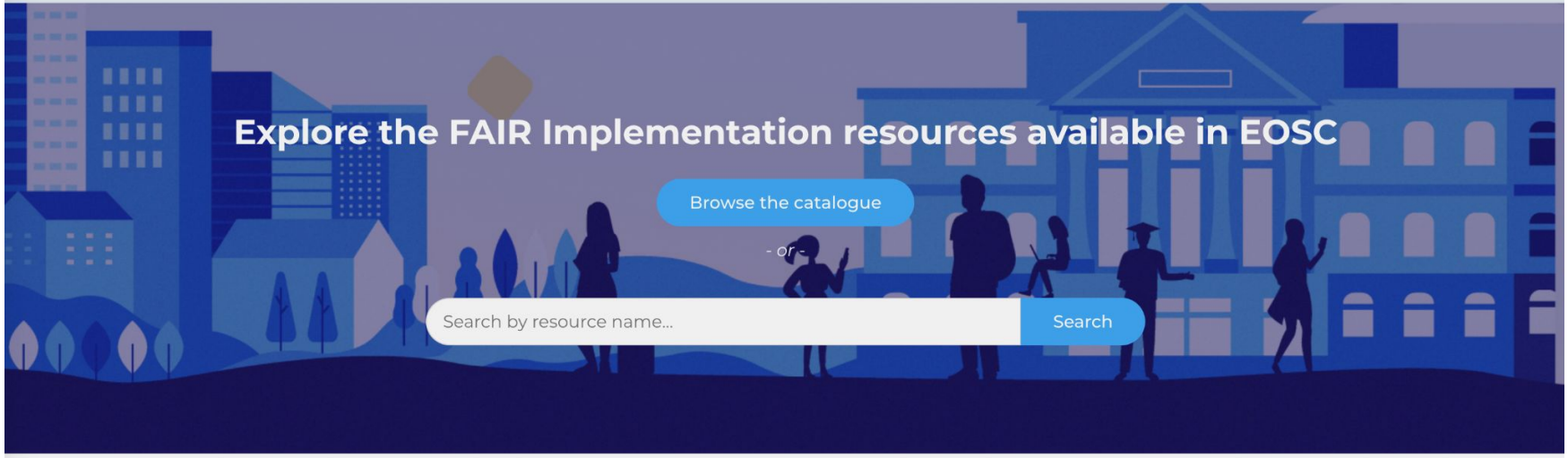
✓ Apply | ✓ Approval | ✓ Approved

Show more

Access to **Data License for AustLit** granted to Dr Peter Sefton

Name	Convict Once
Description	Not Defined
Date Published	1871
@id	arcp://name,AustLit/DerivedMaterial/steconv.xml
Language	English
Conforms To	https://purl.archive.org/language-data-commons/profile#Object
Date Created	1871
Creator	Stephens, J. Brunton (James Brunton) (1835-1902)
Publisher	University of Sydney Library
Citation	Convict Once

5. With a repository to keep data safe and serve it using persistent Identifiers, RO-Crates help make data FAIR



RO-Crate

Research Object Crate (RO-Crate)

RO-Crate has been established as a community effort to practically achieve FAIR packaging of research objects (digital objects like data, methods, software, etc.) with their structured metadata. RO-Crate is based on well-established Web standards and FAIR principles. For its common metadata representations, RO-Crate builds on schema.org, a mature and general mark-up vocabulary used by search engines including Google Dataset Search. RO-Crate libraries are available for Javascript, Python, Ruby and Java, and in addition any RDF tooling supporting JSON-LD can be used (e.g. for knowledge graphs).

Who is the service for

RO-Crate

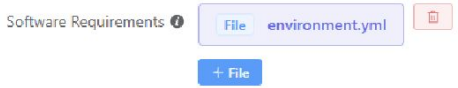
Organisation:
Research Object community

Target users:

The Australian Text Analytics Platform (ATAP)

Binder supports sophisticated Jupyter notebooks with extensive dependencies across all research domains

An ***environment.yml*** file in the repo can specify a Conda environment exactly for reproducibility



6. (bonus point) There are tools which can run a repository from a set of static files on a storage service, in line with the ideas put forward by Prof Suleman at OR 2023

UTS Research Data Portal

Home Help About

Search

Author/s

Hamed Kalhori

Mehrisadat Makki-Alamdari

Bijan Samali

Chul-Woo Kim

Benjamin Halkon

Licence

undefined

2020

< back to search

Ambient Vibration of a Cable-Stayed Bridge

Dataset

This publication is the dataset component of a data paper. A full-scale short-span cable-stayed bridge, located on the top of a wind-exposed hill in the state of the New South Wales (NSW) in Australia, was instrumented to measure its dynamic response to ambient vibrations. The main purpose of the exercise was to generate sufficient ambient vibration datasets necessary for conducting Operational Modal Analysis (OMA). Wind, passing vehicular and pedestrian traffic over the bridge, as well as the vehicles travelling on the highway underneath the bridge provide adequate sources of ambient vibration excitation for this bridge. A dense array of time-synchronised uni-axial accelerometers was permanently mounted on the deck and on the cables of the bridge. Since the structural modal features vary with temperature, the ambient temperature was also continuously recorded. The shear strain response at one end of the bridge was also measured constantly to identify the volume of passing traffic over the bridge. Data acquisition was conducted non-stop for specific periods and the measured data were transferred over a 4G cellular network to the database. It is the intention of the authors that the datasets can be employed for further development and validation of OMA frameworks and will be of interest to the bridge engineering research community.

Ambient Vibration of a Cable-Stayed Bridge

Download all the metadata for Ambient Vibration of a Cable-Stayed Bridge in JSON-LD format

Ambient Vibration of a Cable-Stayed Bridge

@id	./
name	Ambient Vibration of a Cable-Stayed Bridge
@type	Dataset
description	This publication is the dataset component of a data paper. A full-scale short-span cable-stayed bridge, located on the top of a wind-exposed hill in the state of the New South Wales (NSW) in Australia, was instrumented to measure its dynamic response to ambient vibrations. The main purpose of the exercise was to generate sufficient ambient vibration datasets necessary for conducting Operational Modal Analysis (OMA). Wind, passing vehicular and pedestrian traffic over the bridge, as well as the vehicles travelling on the highway underneath the bridge provide adequate sources of ambient vibration excitation for this bridge. A dense array of time-synchronised uni-axial accelerometers was permanently mounted on the deck and on the cables of the bridge. Since the structural modal features vary with temperature, the ambient temperature was also continuously recorded. The shear strain response at one end of the bridge was also measured constantly to identify the volume of passing traffic over the bridge. Data acquisition was conducted non-stop for specific periods and the measured data were transferred over a 4G cellular network to the database. It is the intention of the authors that the datasets can be employed for further development and validation of OMA frameworks and will be of interest to the bridge engineering research community.
funder	
datePublished	2020-07-14T03:00:34.182Z
author	Hamed Kalhori Mehrisadat Makki-Alamdari Bijan Samali Maria Rashidi Benjamin Halkon
identifier	public_ocfl: 0bacede4bd5a39993ec5088e6247cac8
path	./
dateCreated	2020-07-14T03:00:34.182Z
publisher	University of Technology Sydney
hasPart	Bridge_Structural_Drawings Datasets
contactPoint	_:contact/Hamed.Kalhori@uts.edu.au
keywords	Ambient Vibration Dataset; Bridge Structural Analysis; Cable-Stayed Bridge; Operational Modal Analysis.

The Language Data Commons of Australia Data Partnerships (LDaCA-DP), Language Data Commons of Australia Research Data Commons (LDaCA-RDC), and Australian Text Analytics Platform (ATAP) projects received investment (https://doi.org/10.47486/DP768, https://doi.org/10.47486/HIR001, & https://doi.org/10.47486/PL074) from the Australian Research Data Commons (ARDC).

The ARDC is funded by the National Collaborative Research Infrastructure Strategy (NCRIS).

A web version of this presentation with notes can be found at:
https://www.ldaca.edu.au/news/posts/open-repositories-2024-ro-crate