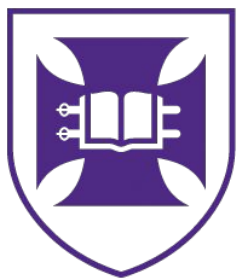


Building domain specific vocabularies for packaging and archiving research and cultural data

Peter Sefton & Mike Lynch



**THE UNIVERSITY
OF QUEENSLAND**
AUSTRALIA



**australian text
analytics platform**
atap.edu.au

CREATE CHANGE



Australia's Academic
and Research Network



MONASH
University



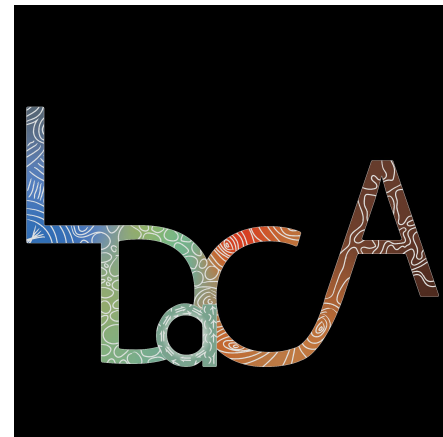
**Australian
National
University**



AUSTRALIAN
ACCESS FEDERATION
AAF.EDU.AU



THE UNIVERSITY OF
SYDNEY



With thanks for their contribution:



AIATSIS



Australian Research Data Commons



The Language Data Commons of Australia (LDaCA) and Australian Text Analytics Platform (ATAP) projects received investment (<https://doi.org/10.47486/DP768> and <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).

Yesterday's keynote speaker **Arofan Gregory** asked:

Why, when the [Linked Data] technology and standards can easily solve the problem, do we remain unable to leverage these critical assets?

Addressable
resources



Local Data



<https://orcid.org/0000-0001-2345-6789>

ID? Title? Description?



Who created this data?



What parts does it have?



When?



What is it about?



How can it be reused?



As part of which project?



Who funded it?



How was it made?



<https://example.com/my-microscope>



CARE and FAIR principles inform all this work, and we have implemented systems, but that's another presentation

5-6789

Addressable resources



Local Data

ID? Title? Description?

 Who created this data?

 What parts does it have?

 When?

 What is it about?

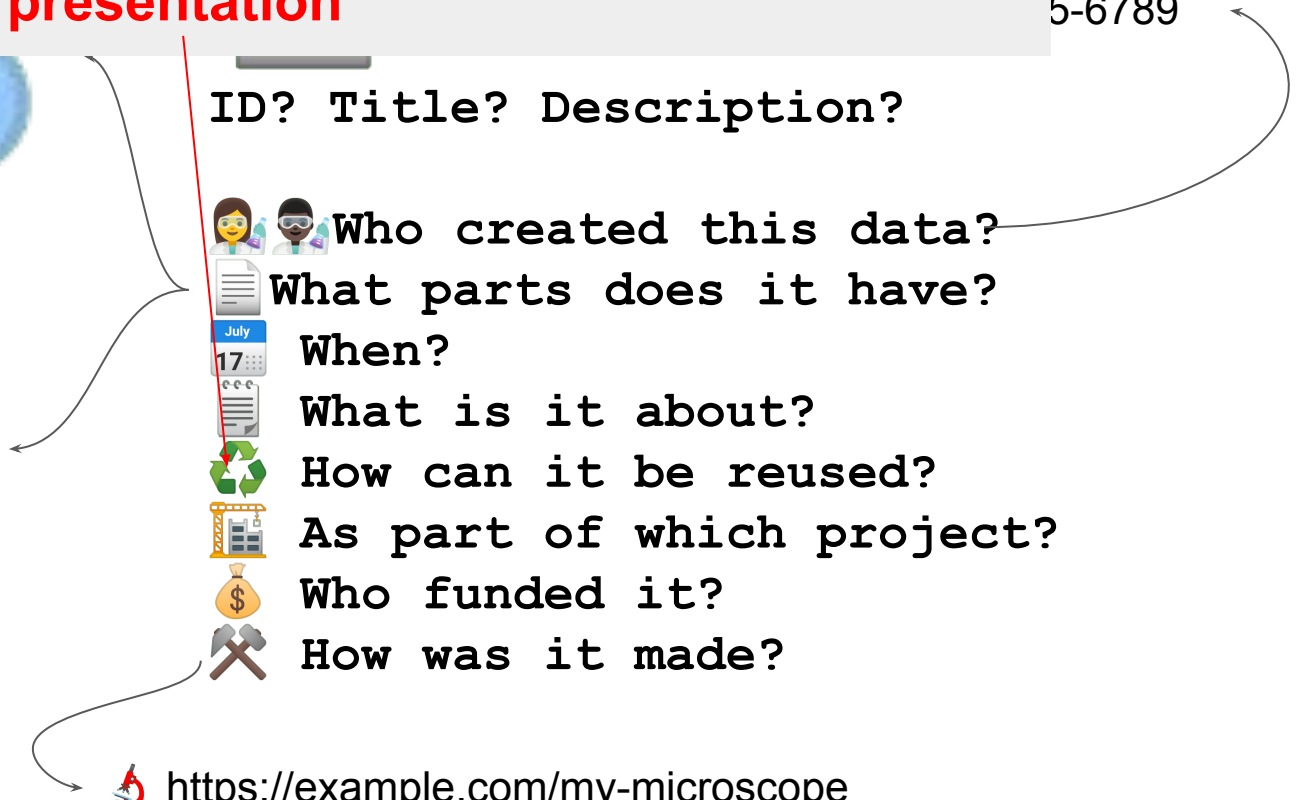
 How can it be reused?

 As part of which project?

 Who funded it?

 How was it made?

 <https://example.com/my-microscope>





```
|-- 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**



A close-up photograph of a black and white puppy looking upwards against a bright, hazy background.

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

📄 webinar-886835...ics ^ Show all ×

eMelbourne - The Encyclopedia of Melbourne Online

[Download all the metadata for eMelbourne - The Encyclopedia of Melbourne Online in JSON-LD format](#)

[Check this crate](#)

Browse files eMelbourne - The Encyclopedia of Melbourne Online

@id	/
name [?]	eMelbourne - The Encyclopedia of Melbourne Online
@type	Dataset
description [?]	eMelbourne is a biographical, bibliographical and archival database about Melbourne, Victoria, Australia, with links to related articles and images.
datePublished [?]	2017-10-26T00:00:00.000Z
keywords [?]	<ul style="list-style-type: none">• Victoria• Melbourne• history• heritage• eMelbourne• encyclopedia• architecture• education• fine arts

eMelbourne - The Encyclopedia of Melbourne Online

[Download all the metadata for eMelbourne - The Encyclopedia of Melbourne Online in JSON-LD format](#)

[Check this crate](#)

Browse files eMelbourne - The Encyclopedia of Melbourne Online

@id	/
name [?]	eMelbourne - The Encyclopedia of Melbourne Online
@type	Dataset
description [?]	eMelbourne is a biographical, bibliographical and archival database about Melbourne, Victoria, Australia, with links to related articles and images.
datePublished [?]	2017-10-26T00:00:00.000Z
keywords [?]	<ul style="list-style-type: none">• Victoria• Melbourne• history• heritage• eMelbourne• encyclopedia• architecture• education• fine arts

eMelbourne - The Encyclopedia of Melbourne Online

[Download all the metadata for eMelbourne - The Encyclopedia of Melbourne Online in JSON-LD format](#)

[Check this crate](#)

[Download: objects/images/ca000187.jpg](#)





Search...

Advanced Search beta

Filters

Collection

1 >

Member Of

1 >

Access

1 >

Record Type

11 >

Language

55 >

Total: 1236 Index entries (Collections, Objects, Files and Notebooks)

RESET SEARCH

Sort by: Collections ▾

Order by: Descending ▾

< 1 2 3 4 5 6 ... 124 >

Opening Australia's Multilingual Archive

Type: Dataset RepositoryCollection

Language: Albanian, Arabic, Armenian, Bengali, Bosnian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, Greek (Modern), Gujarati, Hebrew, Hindi, Hungarian, Italian, Japanese, Khmer, Korean, Lao, Latvian, Lithuanian, Maltese, Nepali, Norwegian, Panjabi; Punjabi, Persian, Portuguese, Russian, Serbian, Slovak, Slovenian, Spanish, Thai, Turkish, Ukrainian, Urdu, Vietnamese, Welsh, Yiddish

Opening the Multilingual Archive of Australia brings together records from national and international collections in languages from around the world. We seek to rethink and enlarge narratives about Australia.

Objects: 1235,

[See more](#)

Voyage en Australie et en Nouvelle-Zélande

Type: Text RepositoryObject

Language: French



Opening Australia's Multilingual Archive > Goa eist Australië terug

Not Defined

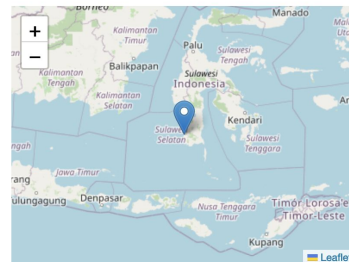
Date Published

Not Defined

@id

#rec_432

_geolocation



_main Collection

arcp://name,omaa-test

_sub Collection

Abstract

The king of Goa has requested to the British Ambassador in Jakarta, that Australia be returned to him since it was formerly part of the kingdom of Makassar, under the name of Marege.

Access

No License

[Default LDaCA No License](#)

Public Metadata and Text is Searchable

Member Of

[Opening Australia's Multilingual Archive](#)

Other Objects in this Collection

[Voyage en Australie et en Nouvelle-Zélande](#)
[Señor : el capitan Pedro Fernandez de Quirós : despues q' puse en practica el descubrimiento de las tierras incognitas del Sur](#)
[The gold digger \[manuscript\] = Der...](#)

Person

A Schema.org Type

Thing > Person

[more...]

A person (alive, dead, undead, or fictional).

Property	Expected Type	Description
Properties from Person		
additionalName	Text	An additional name for a Person, can be used for a middle name.
address	PostalAddress or Text	Physical address of the item.
affiliation	Organization	An organization that this person is affiliated with. For example, a school/university, a club, or a team.
agentInteractionStatistic	InteractionCounter	The number of completed interactions for this entity, in a particular role (the 'agent'), in a particular action (indicated in the statistic), and in a particular context (i.e. interactionService).
alumniOf	EducationalOrganization or Organization	An organization that the person is an alumni of. Inverse property: alumni
award	Text	An award won by or for this item. Supersedes awards.
birthDate	Date	Date of birth.
birthPlace	Place	The place where the person was born.
brand	Brand or Organization	The brand(s) associated with a product or service, or the brand(s) maintained by an organization or business person.
callSign	Text	A callsign, as used in broadcasting and radio communications to identify people, radio and TV stations, or vehicles.
children	Person	A child of the person.
colleague	Person or URL	A colleague of the person. Supersedes colleagues.
contactPoint	ContactPoint	A contact point for a person or organization. Supersedes contactPoints.
deathDate	Date	Date of death.
deathPlace	Place	The place where the person died.
duns	Text	The Dun & Bradstreet DUNS number for identifying an organization or business person.
email	Text	Email address.
familyName	Text	Family name. In the U.S., the last name of a Person.
faxNumber	Text	The fax number.
follows	Person	The most generic uni-directional social relation.
funder	Organization or Person	A person or organization that supports (sponsors) something through some kind of financial contribution.
funding	Grant	A Grant that directly or indirectly provide funding or sponsorship for this item. See also ownershipFundingInfo. Inverse property: fundedItem
gender	GenderType or Text	Gender of something, typically a Person, but possibly also fictional characters, animals, etc. While https://schema.org/Male and https://schema.org/Female may be used, text strings are also acceptable for people who do not identify as a binary gender. The gender property can also be used in an extended sense to cover e.g. the gender of sports teams. As with the gender of individuals, we do not try to enumerate all possibilities. A mixed-gender SportsTeam can be indicated with a text value of "Mixed".
givenName	Text	Given name. In the U.S., the first name of a Person.
globalLocationNumber	Text	The Global Location Number (GLN, sometimes also referred to as International Location Number or ILN) of the respective organization, person, or place. The GLN is a 13-digit number used to identify parties and physical locations.
hasCredential	EducationalOccupationalCredential	A credential awarded to the Person or Organization.
hasOccupation	Occupation	The Person's occupation. For past professions, use Role for expressing dates.
hasOfferCatalog	OfferCatalog	Indicates an OfferCatalog listing for this Organization, Person, or Service.
hasPOS	Place	Points-of-Sales operated by the organization or person.

```
"http://purl.org/dc/terms/identifier": [
  {
    "@type":
"http://www.w3.org/2001/XMLSchema#token",
    "@value": "AAC"
  },

```

Colleagues sent us this

```
w3.org/2001/XMLSchema#token
<xs:restriction base="xs:string">
  <xs:whiteSpace value="replace" id="normalizedString.whiteSpace"/>
</xs:restriction>
</xs:simpleType>
<xs:simpleType name="token" id="token">
  <xs:annotation>
    <xs:documentation source="http://www.w3.org/TR/xmlschema-2/#token"/>
  </xs:annotation>
  <xs:restriction base="xs:normalizedString">
    <xs:whiteSpace value="collapse" id="token.whiteSpace"/>
  </xs:restriction>
</xs:simpleType>
<xs:simpleType name="language" id="language">
  <xs:annotation>
    <xs:documentation source="http://www.w3.org/TR/xmlschema-2/#language"/>
  </xs:annotation>
  <xs:restriction base="xs:token">
    <xs:pattern value="[a-zA-Z]{1,8}(-[a-zA-Z0-9]{1,8})*" id="language.pattern">
      <xs:annotation>
        <xs:documentation source="http://www.ietf.org/rfc/rfc3066.txt"> pattern specifies the content of section 2.12 of XML 1.0e2 and RFC
      </xs:documentation>
    </xs:annotation>
  </xs:pattern>
</xs:restriction>
</xs:simpleType>
<xs:simpleType name="IDREFS" id="IDREFS">
  <xs:annotation>
    <xs:appinfo>
      <hfp:hasFacet name="length"/>
      <hfp:hasFacet name="minLength"/>
      <hfp:hasFacet name="maxLength"/>
      <hfp:hasFacet name="enumeration"/>
      <hfp:hasFacet name="whiteSpace"/>
      <hfp:hasFacet name="pattern"/>
      <hfp:hasProperty name="ordered" value="false"/>
      <hfp:hasProperty name="bounded" value="false"/>
      <hfp:hasProperty name="cardinality" value="countably infinite"/>
      <hfp:hasProperty name="numeric" value="false"/>
    </xs:appinfo>
    <xs:documentation source="http://www.w3.org/TR/xmlschema-2/#IDREFS"/>
  </xs:annotation>
</xs:restriction>
</xs:simpleType>
```

Which resolves to this Page, this might be the token we're looking for

hasFile [?]



- [objects/images/ca000187.jpg](http://pcdm.org/objects/images/ca000187.jpg)
- [objects/thumbs/tn_ca000187.jpg](http://pcdm.org/objects/thumbs/tn_ca000187.jpg)

pcdm.org/2016/04/18/models#hasFile



pcdm:hasFile

URI	http://pcdm.org/models#hasFile
Label	has file
Description	Links to a File contained by this Object.
Type	rdf:Property
Domain	pcdm:Object
Range	pcdm:File

pcdm:hasMember

URI	http://pcdm.org/models#hasMember
Label	has member
Description	Links to a subsidiary Object or Collection. Typically used to link to component parts, such as a book linking to a page. Note on transitivity: hasMember is not defined as transitive, but applications may treat it as transitive as local needs dictate.
Type	rdf:Property
Domain	ore:Aggregation
Range	ore:Aggregation

pcdm:hasRelatedObject

<https://purl.archive.org/language-data-commons/terms>*

Language Data Commons Schema Terms [↗](#)

This is a language data Schema, in the style of the Schema.org schema. It is based on OLAC terms for use in the ATAP and LDaCA projects and is published at <https://purl.archive.org/language-data-commons/terms>. This schema builds on schema.org and is intended to be used with the Language Data Commons RO-Crate profile: <https://purl.archive.org/language-data-commons/profile>

Classes [↗](#)

[Annotation](#) | [CollectionEvent](#) | [CollectionProtocol](#) | [DataDepositLicense](#) | [PrimaryMaterial](#) | [DataReuseLicense](#)

Properties [↗](#)

[accessControlList](#) | [annotationOf](#) | [annotationType](#) | [annotator](#) | [author](#) | [authorizationWorkflow](#) | [channels](#) | [collectionEventType](#) | [collectionProtocolType](#) | [compiler](#) | [consultant](#) | [dataInputter](#) | [depositor](#) | [derivationOf](#) | [developer](#) | [doi](#) | [editor](#) | [geoJSON](#) | [hasAnnotation](#) | [hasCollectionProtocol](#) | [hasDerivation](#) | [illustrator](#) | [indexableText](#) | [interpreter](#) | [interviewee](#) | [interviewer](#) | [isDeidentified](#) | [subjectLanguage](#) | [modality](#) | [openAccessIndex](#) | [participant](#) | [performer](#) | [person](#) | [photographer](#) | [recorder](#) | [register](#) | [researchParticipant](#) | [researcher](#) | [responder](#) | [signer](#) | [singer](#) | [speaker](#) | [sponsor](#) | [transcriber](#) | [translator](#) | [access](#) | [reviewDate](#)

DefinedTerms [↗](#)

[Coded](#) | [Dialogue](#) | [Drama](#) | [ElicitationTask](#) | [Formulaic](#) | [Gesture](#) | [Handwritten](#) | [Informational](#) | [Interview](#) | [Lexicon](#) | [Ludic](#) |

Defined Term: SignedLanguage [↗](#)

The resource contains data for which the medium of interaction was signing.

Is an expected value for the following property [↗](#)

[[communicationMode](#)] |

Operating in "Schema Definition Mode" defining a schema for language resources

Profile: Schema profile

Selected Directory: ontology

Language Data Commons Schema Terms / SignedLanguage

Create New Entity

SignedLanguage

Remove

Links from All Entities Unlinked Entities

signed < 1 >

DefinedTerm SignedLanguage

About

@id → https://purl.archive.org/language-data-commons/terms#SignedLanguage

Unique ID.

Related People, Orgs & Works

Structure

@type → DefinedTerm

+ Select

Provenance

The type of the entity.

Space & Time

Software & Hardware

Others

Name → SignedLanguage

The name of this DefinedTerm

Description → The resource contains data for which the medium of interaction was signing.

A description of the defined Term

Mentions ← Dataset Language Data Commons Schema Terms

Same As → + URL

URL of an equivalent term

```
{
```

```
  "@id": "https://purl.archive.org/language-data-commons/terms#SignedLanguage",
```

```
  "@type": "DefinedTerm",
```

```
  "name": "SignedLanguage",
```

```
  "description": "The resource contains data for which the medium of  
                  interaction was signing.",
```

```
  "inDefinedTermSet": {
```

```
    "@id":
```

```
    "https://purl.archive.org/language-data-commons/terms#CommunicationModeTerms"
```

```
  }
```

```
},
```



USING the schema to drive an editing interface

Open Directory Load Files Bulk Add Save Close

Profile: Language Data Commons top level Collection (corpus)

Selected Directory: **cooee-attachments**

A Corpus of Oz Early English (COOEE)

Create New Entity

A Corpus of Oz Early English (COOEE)

Links from All Entities Ur

About

Related People, Orgs & Works

Structure

Provenance

Space & Time

Software & Hardware

Others

Author → Person Clemens W. A. Fritz

+ ...

The person or organisation responsible for creating this collection of data.

Publisher → + Organization

The organisation responsible for releasing this dataset.

Funder → + Organization

The organisation(s) responsible for funding the creation or collection of this dataset.

COOEE

1 Total 6

Dataset RepositoryCollection
A Corpus of Oz Early English (COOEE)

File
A Corpus of Oz Early English (COOEE)

File COOEE contents.XLS

File COOEE.XLS

File COOEE_contents.xlsx

Search...



[Advanced Search](#) beta

Filters

Collection 7 >

Sub Collection 12 >

Access 4 >

Record Type 12 >

Language 1 >

Total: 9421 Index entries (Collections, Objects, Files and Notebooks)

[RESET SEARCH](#)

Sort by: Relevance

Order by: Descending

< **1** 2 3 4 5 6 ... 943 >

[inheritors-text](#)

Type: File DerivedMaterial

Member of: [AustLit](#)

TXT derivative of the original work, without markup

 Relevance Score: 1

[See more](#)

Login needed



[The Land of the Hibiscus Blossum](#)

Type: RepositoryObject

Language: English

Member of: [AustLit](#)

 Relevance Score: 1

[See more](#)

Login needed



[the-land-of-the-hibiscus-blossum-original](#)

Type: File DerivedMaterial

Login needed





Crate-O?

Status: Alpha release -- this should be useable [now in Chrome-based browsers](#), but there will be bugs - please let report bugs using Github issues in this repository.

Crate-O is a browser-based editor for Research Object Crates ([RO-Crate](#)). RO-Crate is a flexible, developer-friendly approach to linked-data description and packaging . Crate-O is designed to:

- describe files on a user's computer and to add contextual information about those files
- optionally skip the files and describe abstract contextual entities such as in a Cultural Collection or an encyclopaedia
- annotate existing resources elsewhere on the web
- import bulk metadata from an Excel spreadsheet

NOTE: Crate-O is for Google Chrome and related browsers ONLY at this stage as it describes files on the users computer, and saves RO-Crate metadata there. We will be releasing a version that can be deployed as part of a service that accesses online resources directly, which will be compatible with other browsers (see the [\[Roadmap\]](#)).

While the current version of Crate-O is designed for editing self-contained RO-Crates (and works fine with crates containing tens of thousands of entities) - our roadmap includes editing fragments of larger linked-data resources, and integration with Arkisto repositories such as the [Oni](#) repository, data API & search portal.

Crate-O is currently developed by the Language Data Commons of Australia ([LDaCA](#)), under the guidance of Peter Sefton as technical lead. If the tool is adopted in other contexts (we are in talks with a few groups about this) then we aim to establish a steering committee / reference group to help guide development.

Your repository details have been saved. ✕

ro-crate-schema-tools

Public



Edit Pins ▾



Unwatch 2 ▾



Fork 2 ▾



Star 0 ▾



main ▾



2 branches



0 tags

Go to file

Add file ▾

<> Code ▾About ⚙️

Your main branch isn't protected

Protect this branch from force pushing or deletion, or require status checks before merging. [Learn more](#)

Protect this branch



ptsefton Updated for IDN examples

a09378e 3 weeks ago 🕒 19 commits

lib	Added support for multiple root types	last month
output	Added implementation for backfilling schema.org entities in a SOSS	3 months ago
sample	New profile builder	4 months ago
test	Added implementation for backfilling schema.org entities in a SOSS	3 months ago
test_data	Added some example crates for the IDN -- this tests Crate-O's abili...	3 weeks ago
.gitignore	Initial commit	last year
LICENSE	Initial commit	last year

Tools for RO-Crate to be used for creating and distributing Schema.org style ontologies and editor Profiles for Crate-O and other RO-Crate editors.

Readme

GPL-3.0 license

Activity

0 stars

2 watching

2 forks

Report repository

Releases

No releases published

oni Public

Edit Pins Unwatch 3 Fork 1 Star 1

main 6 branches 52 tags

Go to file Add file <> Code

About

new Oni implementation

- Readme GPL-3.0 license Activity 1 star 3 watching 1 fork

Report repository

Releases 1

Develop release 1 Latest on May 12, 2022

Packages

No packages published Publish your first package

Your main branch isn't protected Protect this branch

moisbo fixed missing object id query, sequelize model update, #1 8e7b8d9 on Aug 4 261 commits

Table with 3 columns: File name, Commit message, and Time ago. Rows include .vscode, api, configuration, docs, hooks, scripts, .dockerignore, .editorconfig, .gitignore, Dockerfile, and LICENCE.

“Find a way to keep it simple enough without being too wrong”

Stian Soiland-Reyes (I think) copied from the RO-Crate community meeting minutes