Designing a metadata ecosystem for language research based on Research Object Crate (RO-Crate)

Peter Sefton, Nick Thieberger, Marco La Rosa, Simon Musgrave, River Tae Smith, Moises Sacal Bonequi

ARDC

Australian Research Data Commons

NCRIS

National Research Infrastructure for Australia

An Australian Government Initiative

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).

ARC LIEF LE210100013 (2021-2024) Nyingarn: a platform for primary sources in Australian Indigenous languages

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

CREATE CHANGE

atap
australian text analytics platform
atap.edu.au

Partner Institutions:

aarnet
Australia's Academic and Research Network

Australian National University

THE UNIVERSITY OF MELBOURNE

MONASH University

THE UNIVERSITY OF SYDNEY

LDaCA

With thanks for their contribution: AIATSIS

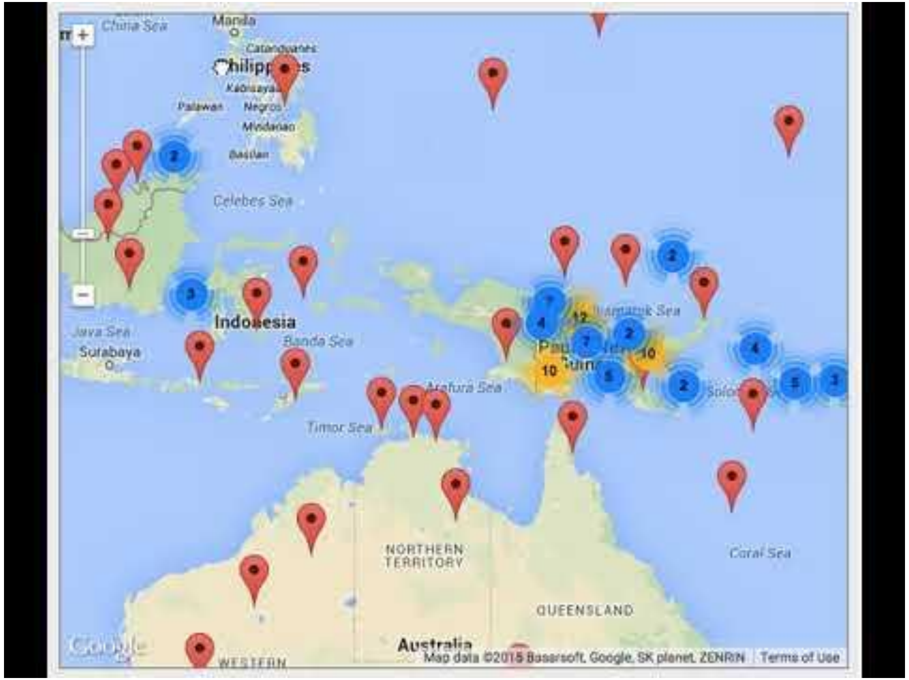
Pacific and Regional Archive for Digital Sources in Endangered Cultures

PARADISEC

Running for 20 years

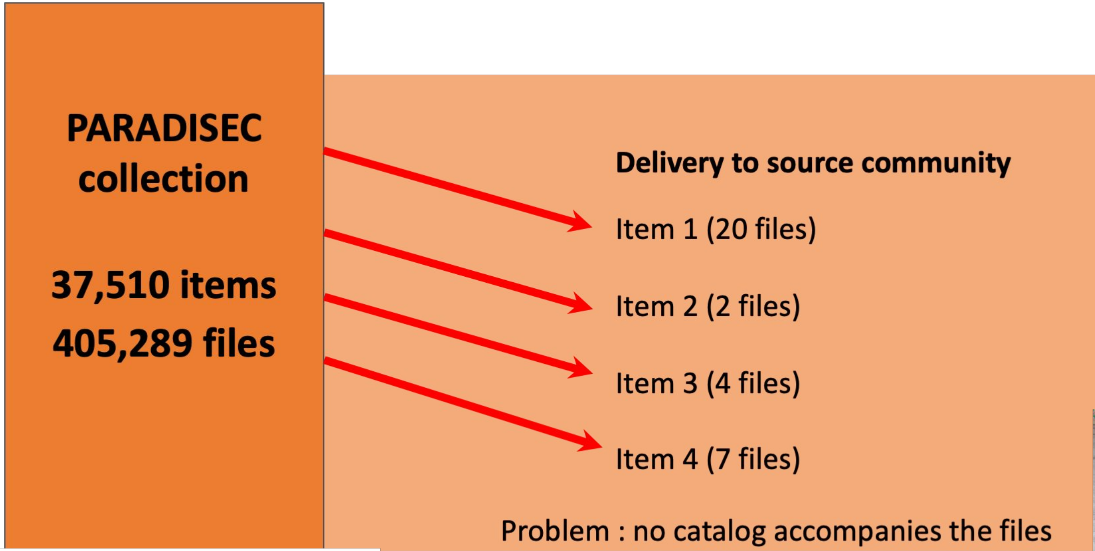
1,337 languages represented
675 collections
37,510 items
405,289 files
15,540 hours (audio)
2,465 hours (video)
193 TB

October 2022



Magnetophonband BASF
Typ LGS 52
DV Cam
Mini DV
VHS-C
NATIONAL STORAGE
Archive Storage Box
ROBINS

PARADISEC ACCESS



PARADISEC collection

37,510 items

405,289 files

Delivery to source community

Item 1 (20 files)

Item 2 (2 files)

Item 3 (4 files)

Item 4 (7 files)

Problem : no catalog accompanies the files

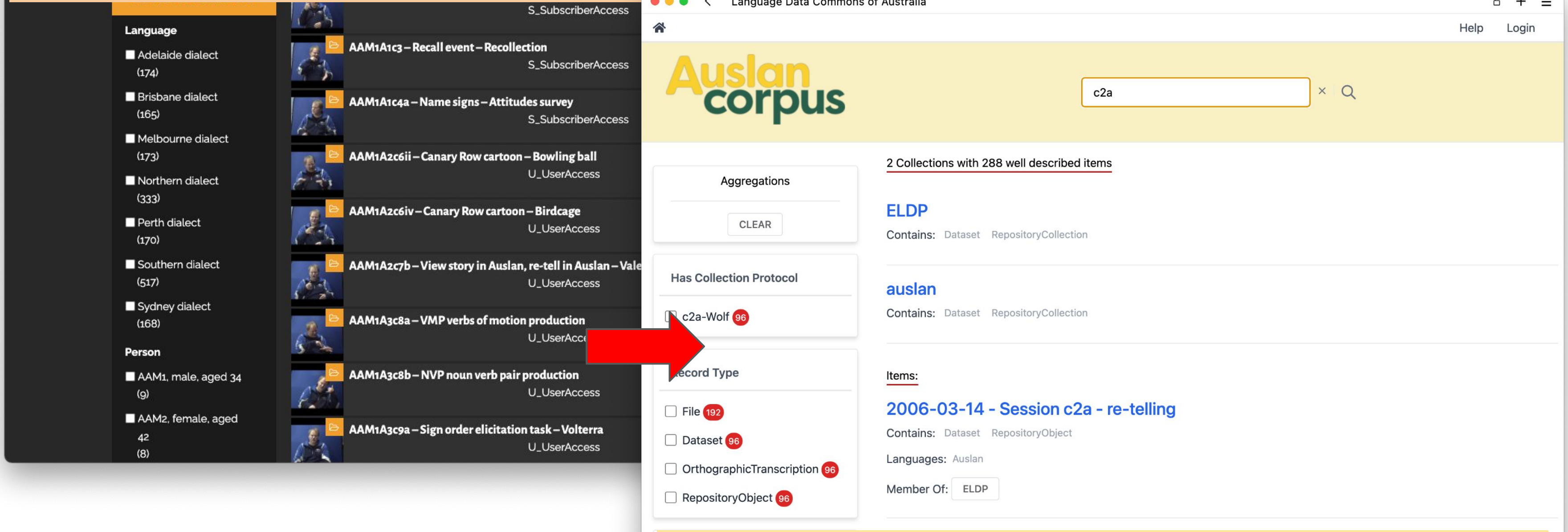
Kalonuk Albert & Gino Joseph in Erakor village

Elisa Alphonse in Erakor village

https://language-archives.services/about/data-loader

AUSLAN CORPUS ACCESS

ELAR: limited search capability, non-standard metadata schema, no ability to index annotation files, no bulk download



Language

- ☐ Adelaide dialect (174)
- ☐ Brisbane dialect (165)
- ☐ Melbourne dialect (173)
- ☐ Northern dialect (333)
- ☐ Perth dialect (170)
- ☐ Southern dialect (517)
- ☐ Sydney dialect (168)

Person

- ☐ AAM1, male, aged 34 (9)
- ☐ AAM2, female, aged 42 (8)

S_SubscriberAccess

AAM1A1c3 – Recall event – Recollection
S_SubscriberAccess

AAM1A1c4a – Name signs – Attitudes survey
S_SubscriberAccess

AAM1A2c6ii – Canary Row cartoon – Bowling ball
U_UserAccess

AAM1A2c6iv – Canary Row cartoon – Birdcage
U_UserAccess

AAM1A2c7b – View story in Auslan, re-tell in Auslan – Vale
U_UserAccess

AAM1A3c8a – VMP verbs of motion production
U_UserAcce

AAM1A3c8b – NVP noun verb pair production
U_UserAccess

AAM1A3c9a – Sign order elicitation task – Volterra
U_UserAccess

Language Data Commons of Australia

Help Login

Auslan corpus

c2a

2 Collections with 288 well described items

Aggregations

CLEAR

Has Collection Protocol

- ☐ c2a-Wolf 96

ecord Type

- ☐ File 192
- ☐ Dataset 96
- ☐ OrthographicTranscription 96
- ☐ RepositoryObject 96

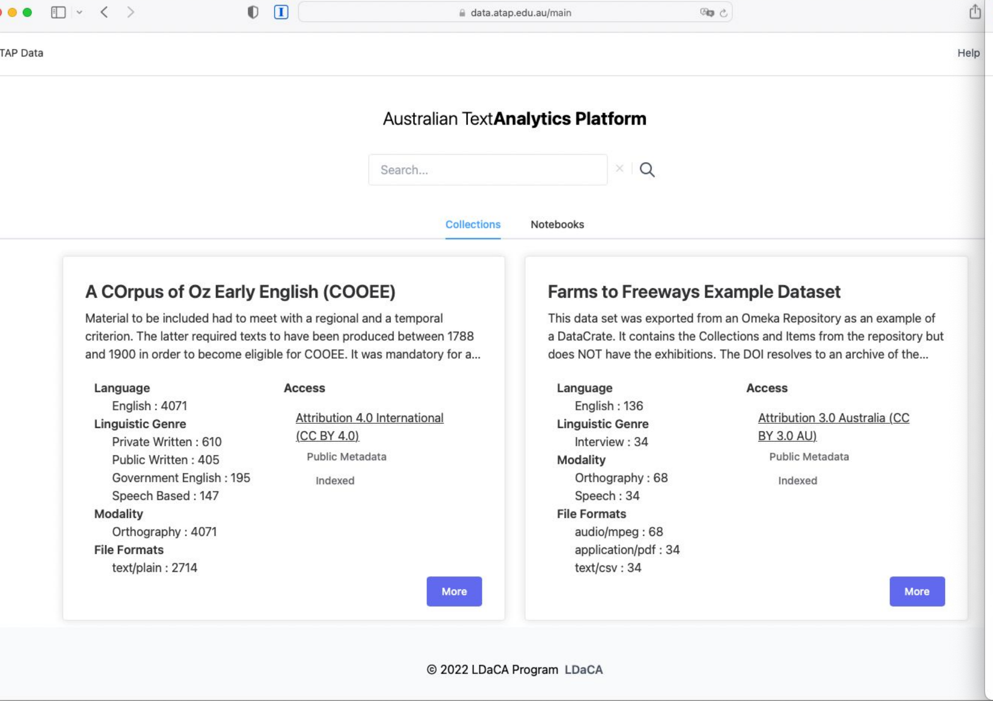
ELDP
Contains: Dataset RepositoryCollection

auslan
Contains: Dataset RepositoryCollection

Items:

2006-03-14 - Session c2a - re-telling
Contains: Dataset RepositoryObject
Languages: Auslan
Member Of: ELDP

LDaCA: rich metadata-first search, portable RO-Crate metadata, indexed annotations, bulk downloading of search results



Australian Text**Analytics Platform**

A COrpus of Oz Early English (COOEE)

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...

Language
English : 4071
Linguistic Genre
Private Written : 610
Public Written : 405
Government English : 195
Speech Based : 147
Modality
Orthography : 4071
File Formats
text/plain : 2714

Access
Attribution 4.0 International (CC BY 4.0)
Public Metadata
Indexed

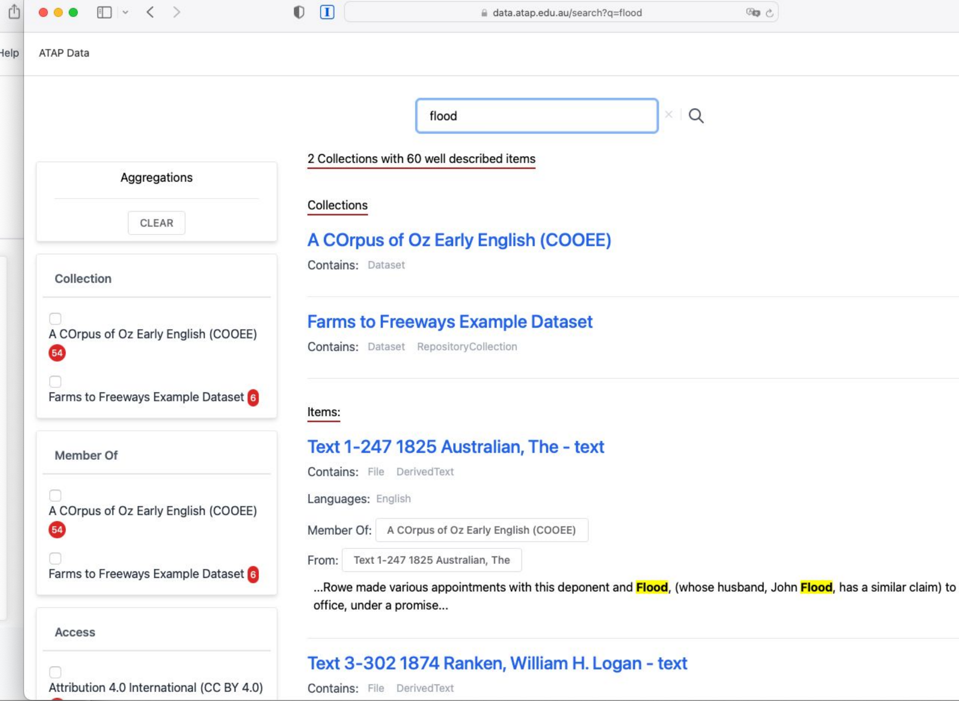
Farms to Freeways Example Dataset

This data set was exported from an Omeka Repository as an example of a DataCrate. It contains the Collections and Items from the repository but does NOT have the exhibitions. The DOI resolves to an archive of the...

Language
English : 136
Linguistic Genre
Interview : 34
Modality
Orthography : 68
Speech : 34
File Formats
audio/mpeg : 68
application/pdf : 34
text/csv : 34

Access
Attribution 3.0 Australia (CC BY 3.0 AU)
Public Metadata
Indexed

© 2022 LDaCA Program LDaCA



2 Collections with 60 well described items

Collection
□ A COrpus of Oz Early English (COOEE) 54
□ Farms to Freeways Example Dataset 6

Member Of
□ A COrpus of Oz Early English (COOEE) 54
□ Farms to Freeways Example Dataset 6

Access
□ Attribution 4.0 International (CC BY 4.0)

Collections

A COrpus of Oz Early English (COOEE)
Contains: Dataset

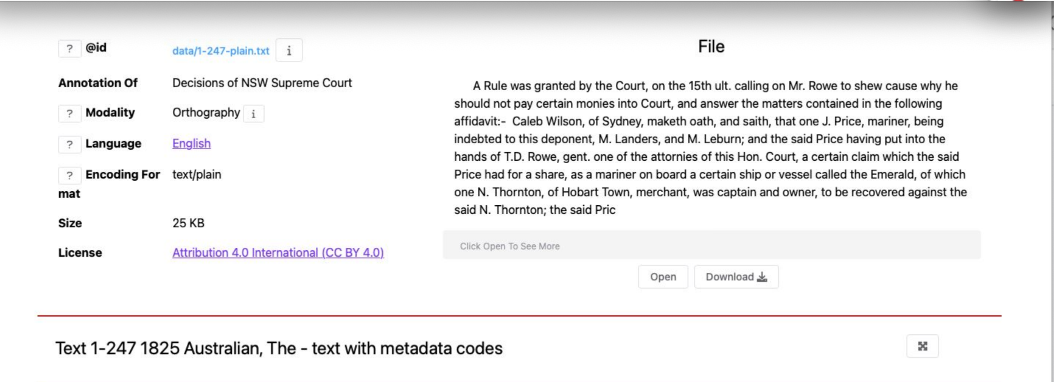
Farms to Freeways Example Dataset
Contains: Dataset RepositoryCollection

Items:

Text 1-247 1825 Australian, The - text
Contains: File DerivedText
Languages: English
Member Of: A COrpus of Oz Early English (COOEE)
From: Text 1-247 1825 Australian, The

...Rowe made various appointments with this deponent and **Flood**, (whose husband, John **Flood**, has a similar claim) to office, under a promise...

Text 3-302 1874 Ranken, William H. Logan - text
Contains: File DerivedText

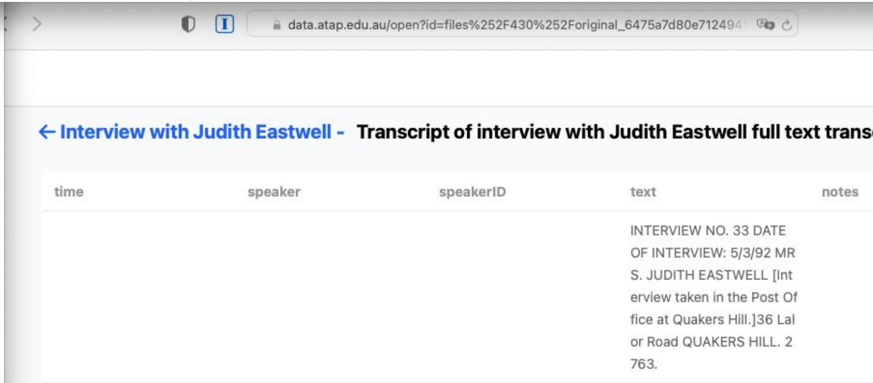


? **@id**	data/1-247-plain.txt i
Annotation Of	Decisions of NSW Supreme Court
? **Modality**	Orthography i
? **Language**	English
? **Encoding Format**	text/plain
Size	25 KB
License	Attribution 4.0 International (CC BY 4.0)

File

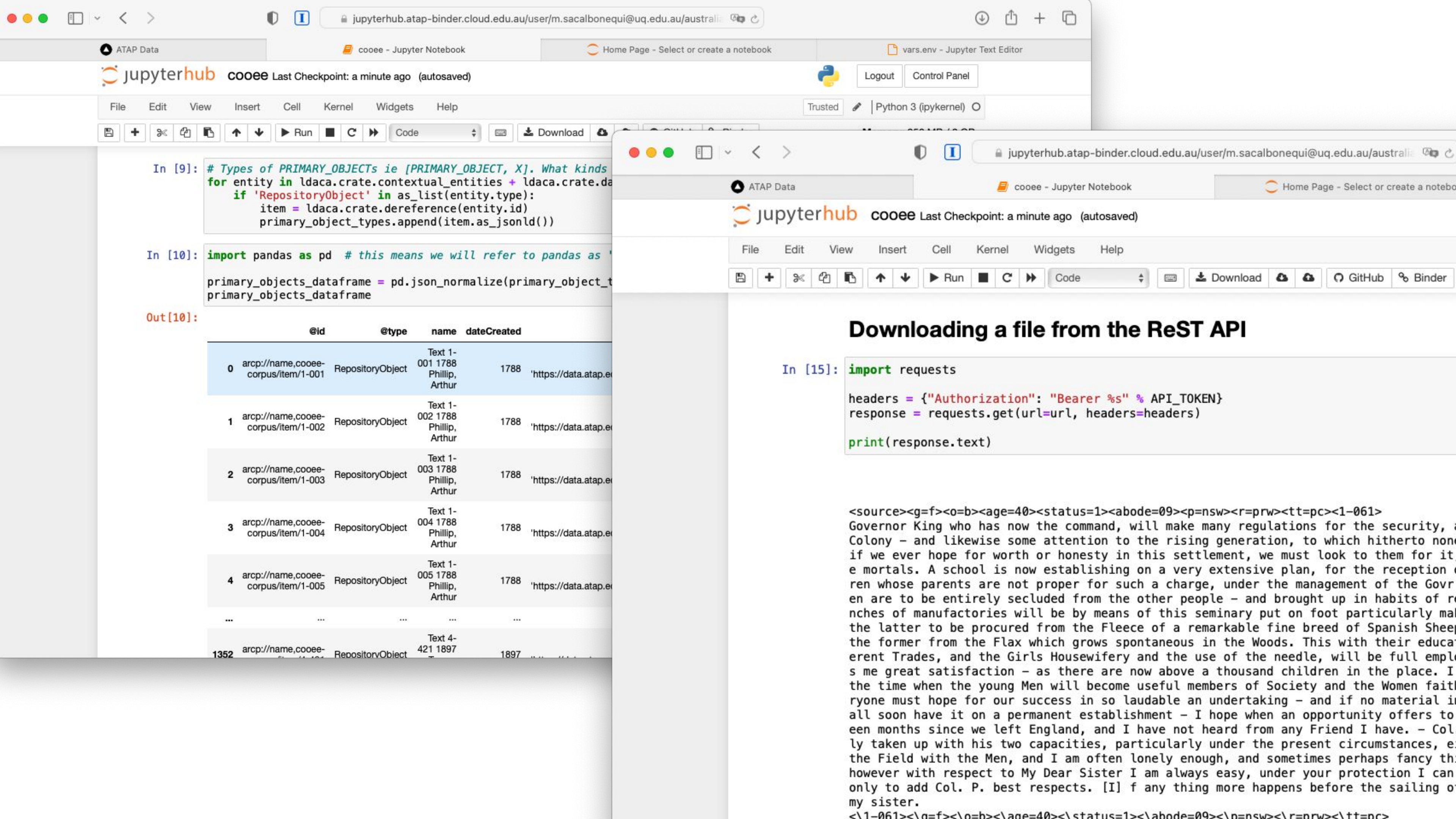
A Rule was granted by the Court, on the 15th ult. calling on Mr. Rowe to shew cause why he should not pay certain monies into Court, and answer the matters contained in the following affidavit:- Caleb Wilson, of Sydney, maketh oath, and saith, that one J. Price, mariner, being indebted to this deponent, M. Landers, and M. Leburn; and the said Price having put into the hands of T.D. Rowe, gent. one of the attornies of this Hon. Court, a certain claim which the said Price had for a share, as a mariner on board a certain ship or vessel called the Emerald, of which one N. Thornton, of Hobart Town, merchant, was captain and owner, to be recovered against the said N. Thornton; the said Pric

Text 1-247 1825 Australian, The - text with metadata codes


data.atap.edu.au/open?id=files%252F430%252Foriginal_6475a7d80e712494

← Interview with Judith Eastwell - Transcript of interview with Judith Eastwell full text trans

time	speaker	speakerID	text	notes
			INTERVIEW NO. 33 DATE OF INTERVIEW: 5/3/92 MR S. JUDITH EASTWELL [Int erview taken in the Post Of fice at Quakers Hill.]36 Lal or Road QUAKERS HILL. 2 763.	


jupyterhub.atap-binder.cloud.edu.au/user/m.sacalbonequi@uq.edu.au/australi
ATAP Data
cooee - Jupyter Notebook
Home Page - Select or create a notebook
vars.env - Jupyter Text Editor
jupyterhub
cooee Last Checkpoint: a minute ago (autosaved)
Logout
Control Panel
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Download
GitHub
Binder

In [9]:

```
# Types of PRIMARY_OBJECTs ie [PRIMARY_OBJECT, X]. What kinds
for entity in ldaca.crate.contextual_entities + ldaca.crate.da
    if 'RepositoryObject' in as_list(entity.type):
        item = ldaca.crate.dereference(entity.id)
        primary_object_types.append(item.as_jsonld())
```

In [10]:

```
import pandas as pd  # this means we will refer to pandas as '

primary_objects_dataframe = pd.json_normalize(primary_object_t
primary_objects_dataframe
```

Out[10]:

	@id	@type	name	dateCreated	
0	arcp://name,cooee-corpus/item/1-001	RepositoryObject	Text 1-001 1788 Phillip, Arthur	1788	'https://data.atap.e
1	arcp://name,cooee-corpus/item/1-002	RepositoryObject	Text 1-002 1788 Phillip, Arthur	1788	'https://data.atap.e
2	arcp://name,cooee-corpus/item/1-003	RepositoryObject	Text 1-003 1788 Phillip, Arthur	1788	'https://data.atap.e
3	arcp://name,cooee-corpus/item/1-004	RepositoryObject	Text 1-004 1788 Phillip, Arthur	1788	'https://data.atap.e
4	arcp://name,cooee-corpus/item/1-005	RepositoryObject	Text 1-005 1788 Phillip, Arthur	1788	'https://data.atap.e
...	...	...	...	...	
1352	arcp://name,cooee-	RepositoryObject	Text 4-421 1897	1897	

Downloading a file from the ReST API

In [15]:

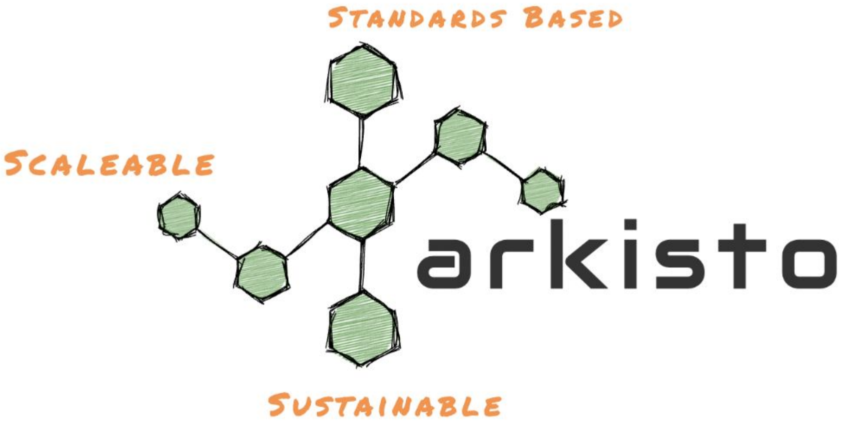
```
import requests

headers = {"Authorization": "Bearer %s" % API_TOKEN}
response = requests.get(url=url, headers=headers)

print(response.text)
```

```
<source><g=f><o=b><age=40><status=1><abode=09><p=nsw><r=prw><tt=pc><1-061>
Governor King who has now the command, will make many regulations for the security,
Colony - and likewise some attention to the rising generation, to which hitherto non
if we ever hope for worth or honesty in this settlement, we must look to them for it
e mortals. A school is now establishing on a very extensive plan, for the reception
ren whose parents are not proper for such a charge, under the management of the Govr
en are to be entirely secluded from the other people - and brought up in habits of r
nches of manufactories will be by means of this seminary put on foot particularly ma
the latter to be procured from the Fleece of a remarkable fine breed of Spanish Shee
the former from the Flax which grows spontaneous in the Woods. This with their educa
erent Trades, and the Girls Housewifery and the use of the needle, will be full empl
s me great satisfaction - as there are now above a thousand children in the place. I
the time when the young Men will become useful members of Society and the Women fait
ryone must hope for our success in so laudable an undertaking - and if no material i
all soon have it on a permanent establishment - I hope when an opportunity offers to
een months since we left England, and I have not heard from any Friend I have. - Col
ly taken up with his two capacities, particularly under the present circumstances, e
the Field with the Men, and I am often lonely enough, and sometimes perhaps fancy th
however with respect to My Dear Sister I am always easy, under your protection I can
only to add Col. P. best respects. [I] f any thing more happens before the sailing o
my sister.
<\1-061><\g=f><\o=b><\age=40><\status=1><\abode=09><\p=nsw><\r=prw><\tt=pc>
```

```
./{collection id}/{item id}
|    ├── {metadata.xml}
|    ├── file1.wav
|    ├── file2.wav
|    ├── file3.mp3
|    ├── ... etc ...
```



A scaleable, standards based platform
for sustainable data.

The basis of Arkisto is that the long-term preservability of well-described data is *always* the first consideration.

Data on an Arkisto deployment is alway available on disc (or object storage) with a complete description *independently* of any services such as websites or APIs. Once the data is safe and well described, Arkisto has a flexible model for how data can be accessed using a variety of services.

Arkisto is built on top of Research Object Crate (RO-Crate) and the Oxford Common File System Layout (OCFL).

With Arkisto there is no messy data migration.

```
{
   "conformsTo": "http://purl.archive.org/language-data-commons/profile"
}
```

Addressable resources

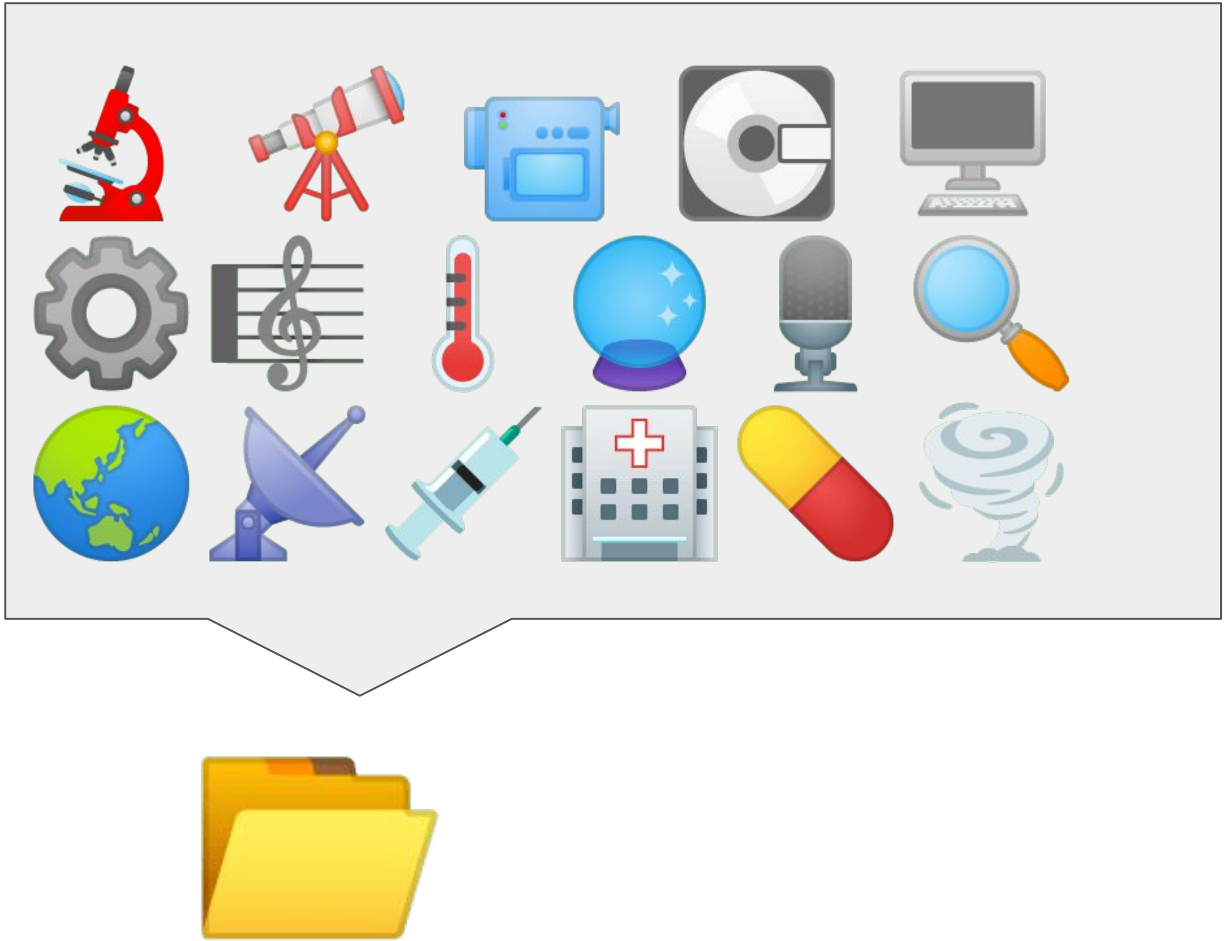
📄 👩‍🔬 https://orcid.org/0000-0001-2345-6789

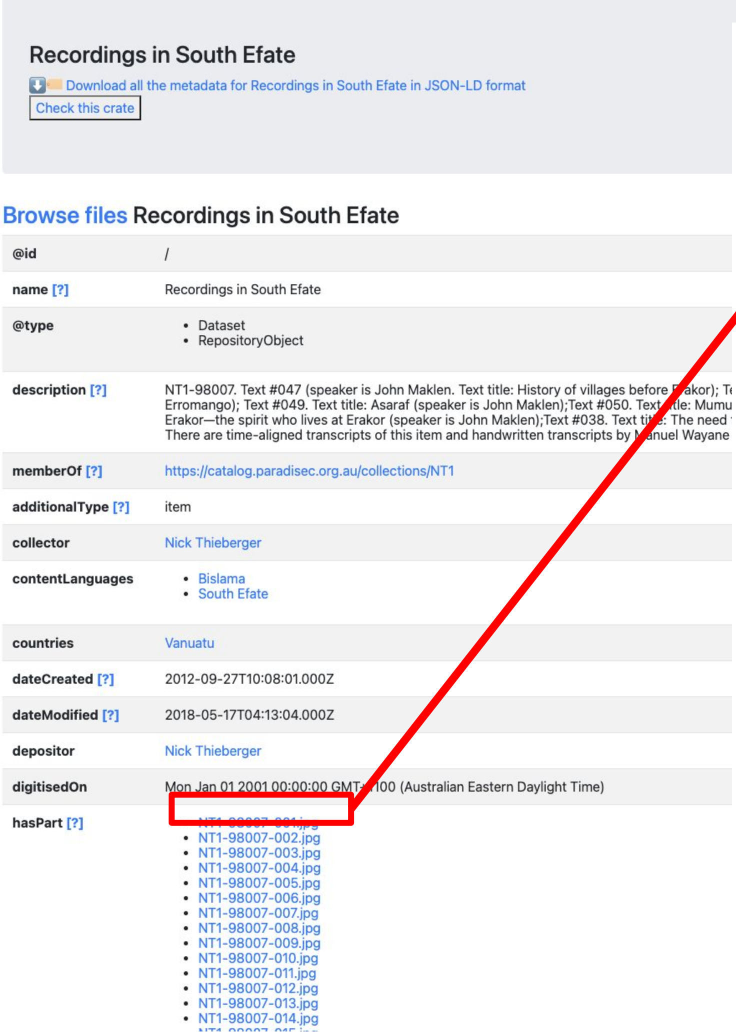
ID? Title? Description?

👩‍🔬👨‍🔬 **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?**

Local Data

🔬 https://en.wikipedia.org/wiki/Scanning_electron_microscope





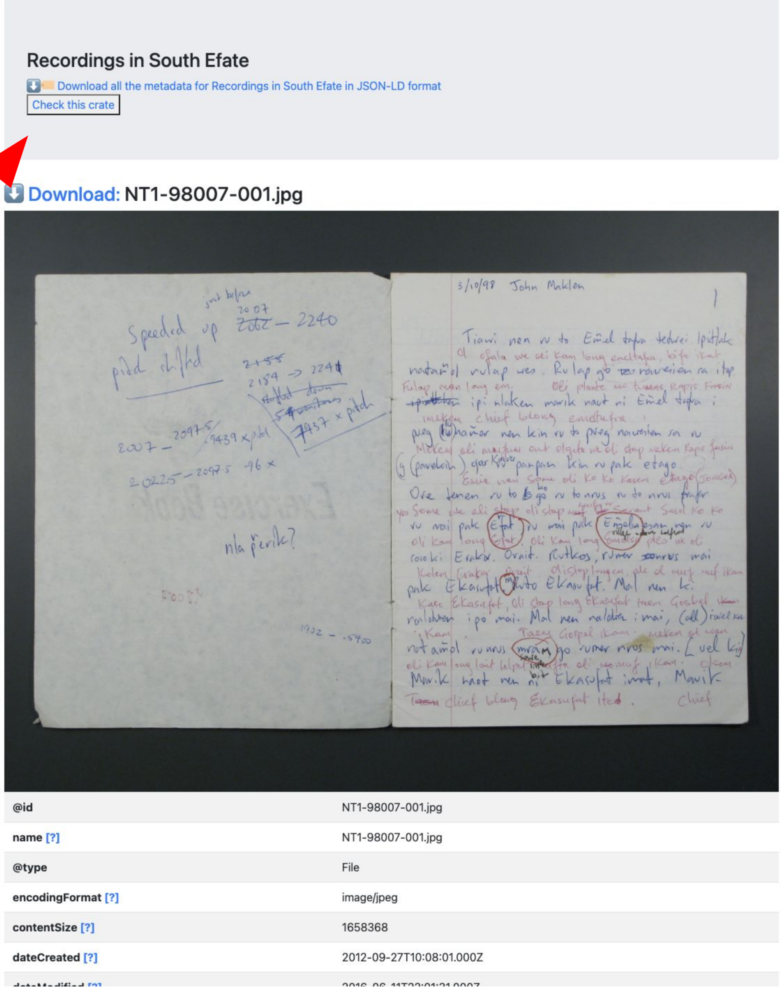
Recordings in South Efate

Download all the metadata for Recordings in South Efate in JSON-LD format

Check this crate

Browse files Recordings in South Efate

@id	/
name [?]	Recordings in South Efate
@type	• Dataset • RepositoryObject
description [?]	NT1-98007. Text #047 (speaker is John Maklen. Text title: History of villages before Erakor); Te Erromango); Text #049. Text title: Asaraf (speaker is John Maklen);Text #050. Text title: Mumu Erakor—the spirit who lives at Erakor (speaker is John Maklen);Text #038. Text title: The need There are time-aligned transcripts of this item and handwritten transcripts by Manuel Wayane
memberOf [?]	https://catalog.paradisec.org.au/collections/NT1
additionalType [?]	item
collector	Nick Thieberger
contentLanguages	• Bislama • South Efate
countries	Vanuatu
dateCreated [?]	2012-09-27T10:08:01.000Z
dateModified [?]	2018-05-17T04:13:04.000Z
depositor	Nick Thieberger
digitisedOn	Mon Jan 01 2001 00:00:00 GMT+1100 (Australian Eastern Daylight Time)
hasPart [?]	• NT1-98007-001.jpg • NT1-98007-002.jpg • NT1-98007-003.jpg • NT1-98007-004.jpg • NT1-98007-005.jpg • NT1-98007-006.jpg • NT1-98007-007.jpg • NT1-98007-008.jpg • NT1-98007-009.jpg • NT1-98007-010.jpg • NT1-98007-011.jpg • NT1-98007-012.jpg • NT1-98007-013.jpg • NT1-98007-014.jpg • NT1-98007-015.jpg



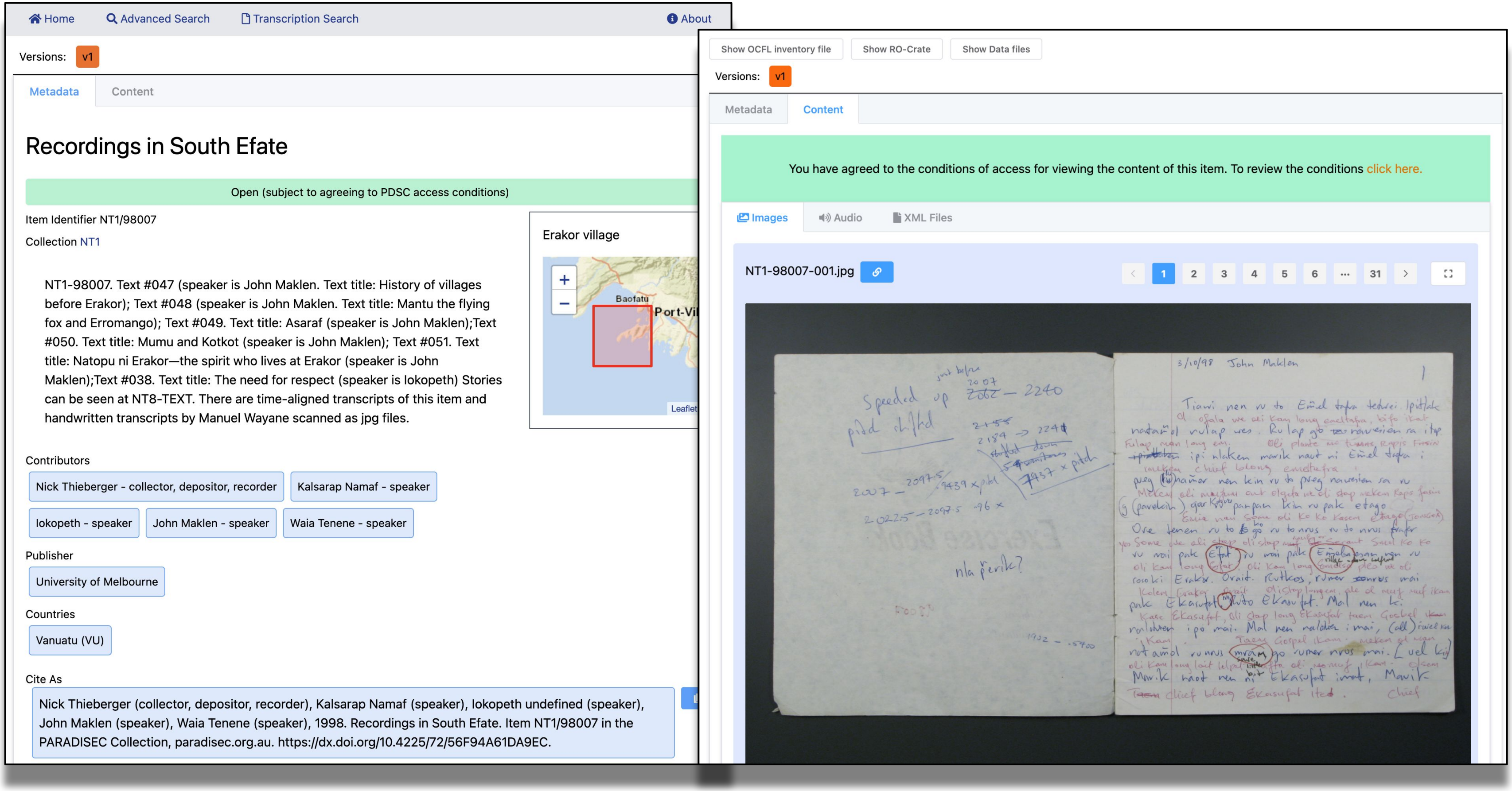
Recordings in South Efate

Download all the metadata for Recordings in South Efate in JSON-LD format

Check this crate

Download: NT1-98007-001.jpg

@id	NT1-98007-001.jpg
name [?]	NT1-98007-001.jpg
@type	File
encodingFormat [?]	image/jpeg
contentSize [?]	1658368
dateCreated [?]	2012-09-27T10:08:01.000Z
dateModified [?]	2016-06-11T22:01:21.000Z



Recordings in South Efate

Open (subject to agreeing to PDSC access conditions)

Item Identifier NT1/98007

Collection NT1

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 Iokopeth) Stories can be seen at NT8-TEXT. There are time-aligned transcripts of this item and handwritten transcripts by Manuel Wayane scanned as jpg files.

Contributors

Nick Thieberger - collector, depositor, recorder

Kalsarap Namaf - speaker

Iokopeth - speaker

John Maklen - speaker

Waia Tenene - speaker

Publisher

University of Melbourne

Countries

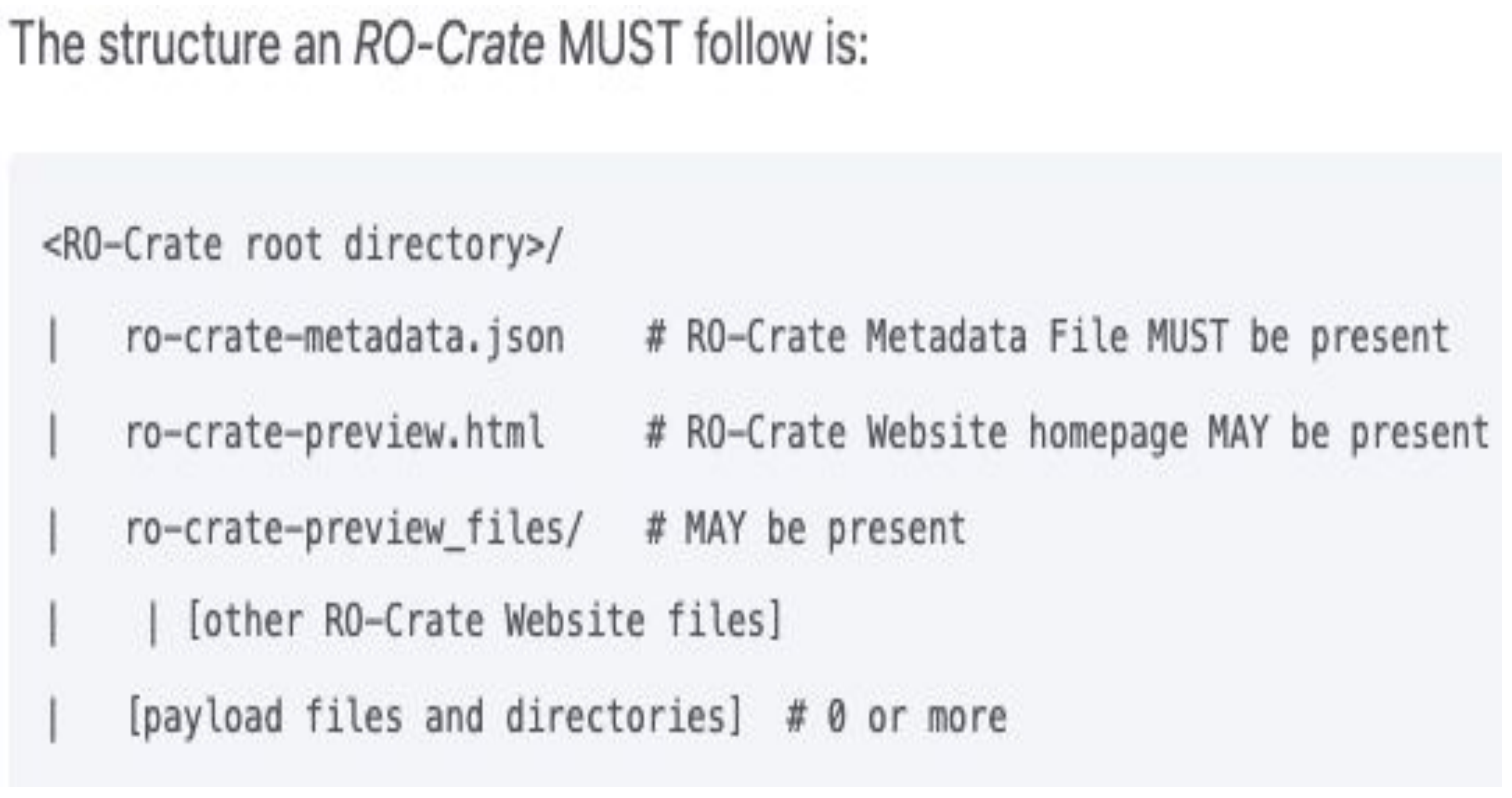
Vanuatu (VU)

Cite As

Nick Thieberger (collector, depositor, recorder), Kalsarap Namaf (speaker), Iokopeth undefined (speaker), John Maklen (speaker), Waia Tenene (speaker), 1998. Recordings in South Efate. Item NT1/98007 in the PARADISEC Collection, paradisec.org.au. https://dx.doi.org/10.4225/72/56F94A61DA9EC.

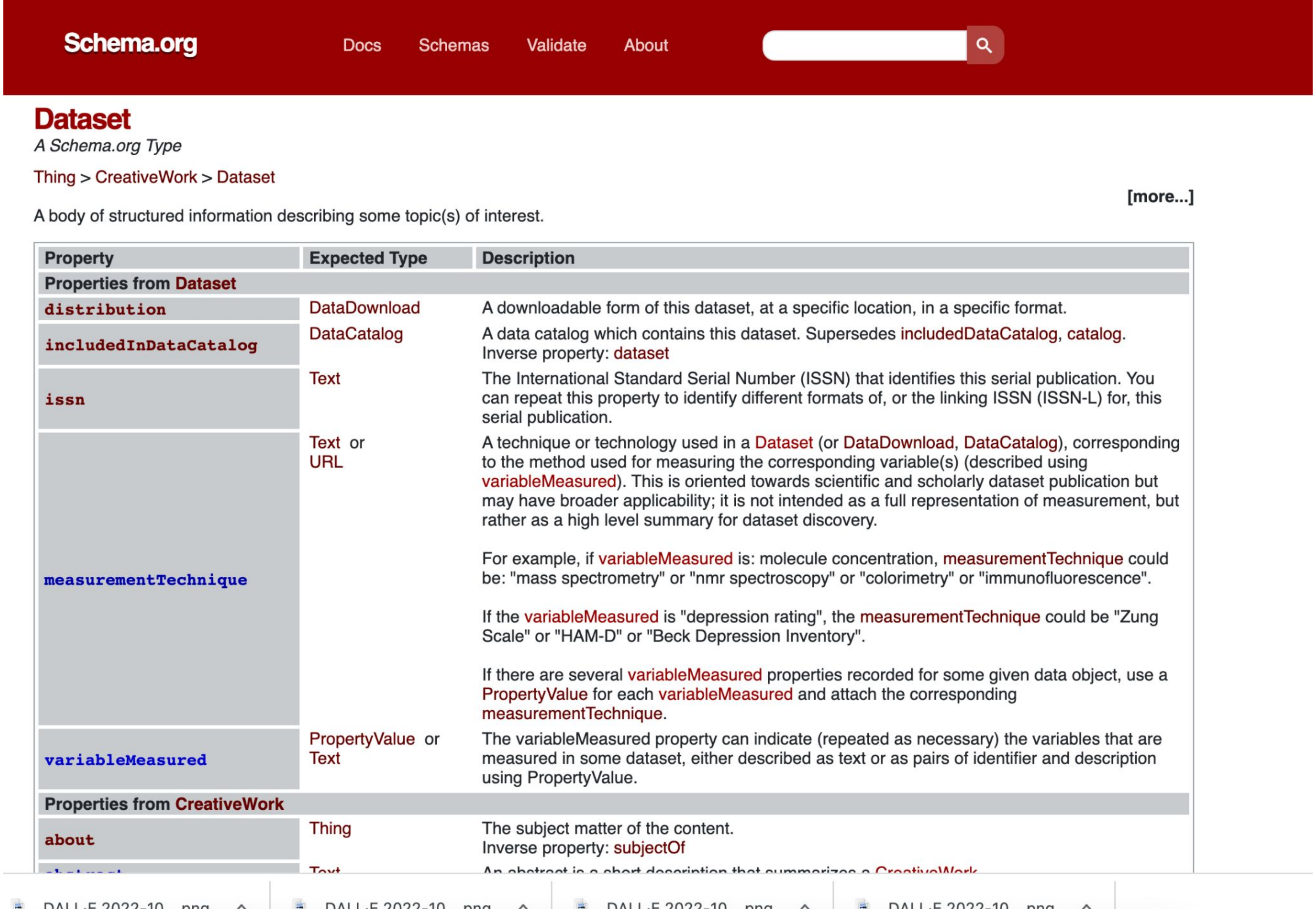
You have agreed to the conditions of access for viewing the content of this item. To review the conditions click here.

https://mod.paradisec.org.au



The structure an *RO-Crate* MUST follow is:

```
<RO-Crate root directory>/
|   ro-crate-metadata.json    # RO-Crate Metadata File MUST be present
|   ro-crate-preview.html     # RO-Crate Website homepage MAY be present
|   ro-crate-preview_files/   # MAY be present
|    | [other RO-Crate Website files]
|   [payload files and directories]  # 0 or more
```



Dataset

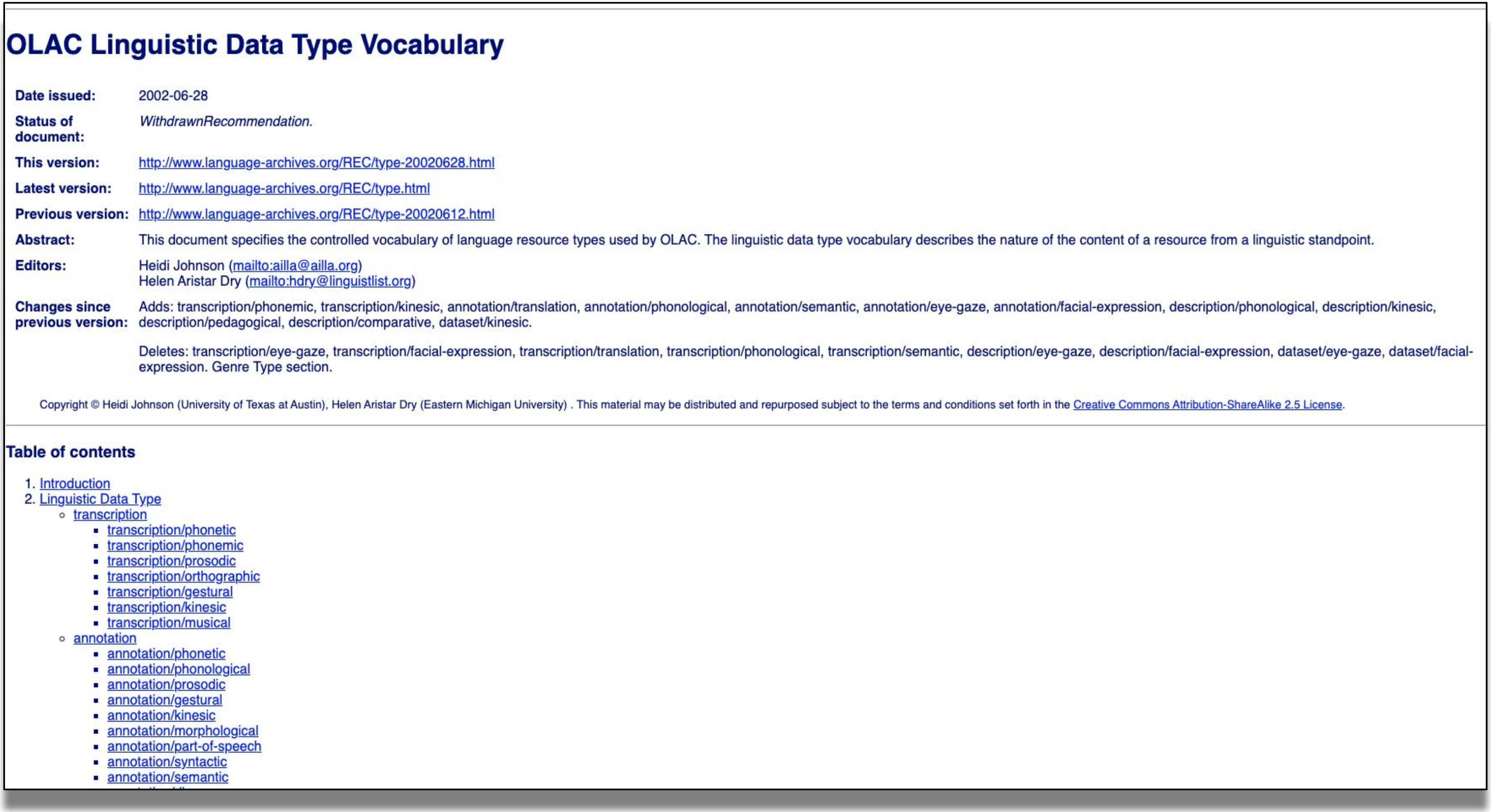
A Schema.org Type

Thing > CreativeWork > Dataset

[more...]

A body of structured information describing some topic(s) of interest.

Property	Expected Type	Description
Properties from Dataset		
`distribution`	DataDownload	A downloadable form of this dataset, at a specific location, in a specific format.
`includedInDataCatalog`	DataCatalog	A data catalog which contains this dataset. Supersedes includedDataCatalog, catalog. Inverse property: dataset
`issn`	Text	The International Standard Serial Number (ISSN) that identifies this serial publication. You can repeat this property to identify different formats of, or the linking ISSN (ISSN-L) for, this serial publication.
`measurementTechnique`	Text or URL	A technique or technology used in a Dataset (or DataDownload, DataCatalog), corresponding to the method used for measuring the corresponding variable(s) (described using variableMeasured). This is oriented towards scientific and scholarly dataset publication but may have broader applicability; it is not intended as a full representation of measurement, but rather as a high level summary for dataset discovery. For example, if variableMeasured is: molecule concentration, measurementTechnique could be: "mass spectrometry" or "nmr spectroscopy" or "colorimetry" or "immunofluorescence". If the variableMeasured is "depression rating", the measurementTechnique could be "Zung Scale" or "HAM-D" or "Beck Depression Inventory". If there are several variableMeasured properties recorded for some given data object, use a PropertyValue for each variableMeasured and attach the corresponding measurementTechnique.
`variableMeasured`	PropertyValue or Text	The variableMeasured property can indicate (repeated as necessary) the variables that are measured in some dataset, either described as text or as pairs of identifier and description using PropertyValue.
Properties from CreativeWork		
`about`	Thing	The subject matter of the content. Inverse property: subjectOf
`abstract`	Text	An abstract is a short description that summarizes a CreativeWork



OLAC Linguistic Data Type Vocabulary

Date issued: 2002-06-28

Status of document: *WithdrawnRecommendation.*

This version: http://www.language-archives.org/REC/type-20020628.html

Latest version: http://www.language-archives.org/REC/type.html

Previous version: http://www.language-archives.org/REC/type-20020612.html

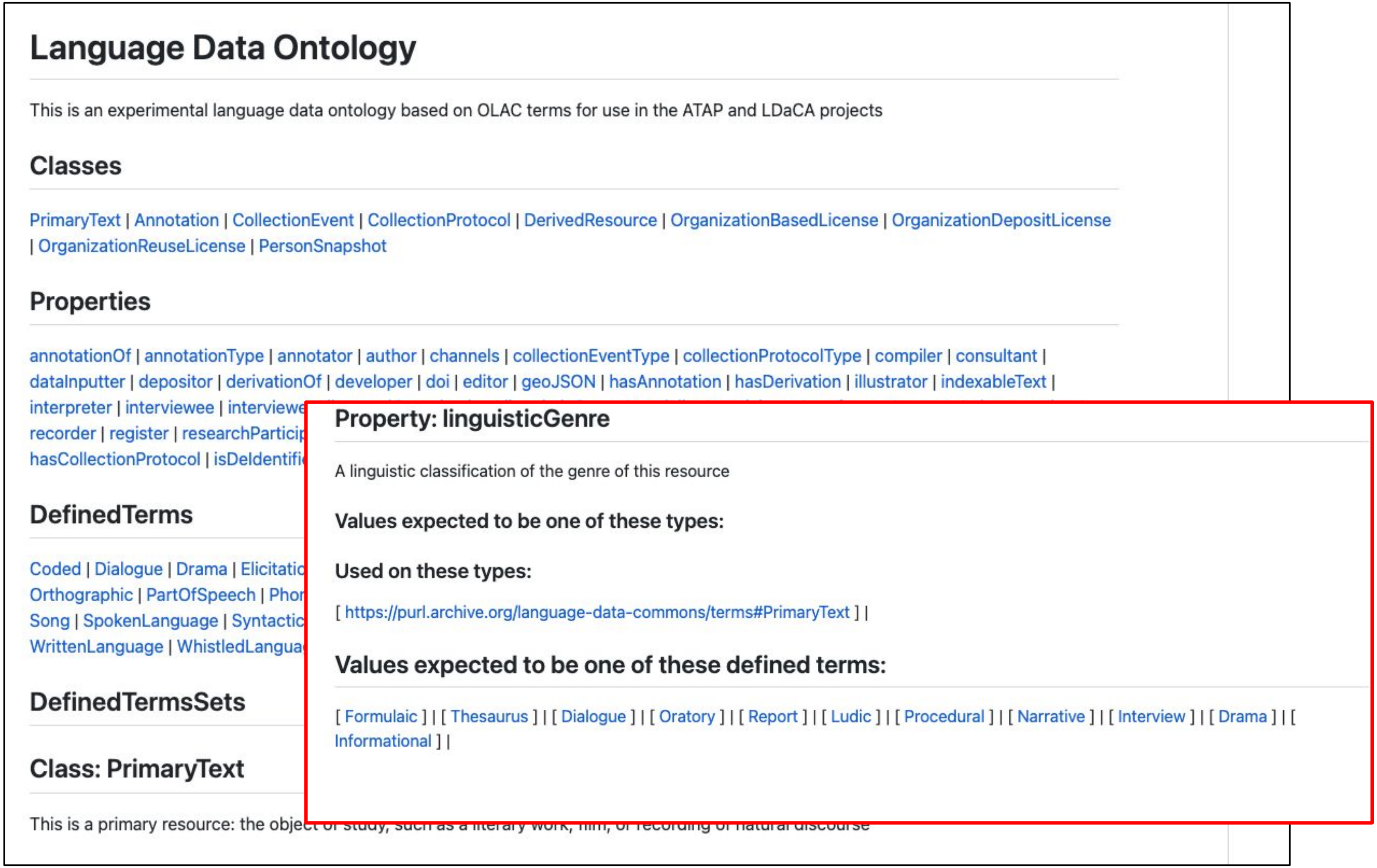
Abstract: This document specifies the controlled vocabulary of language resource types used by OLAC. The linguistic data type vocabulary describes the nature of the content of a resource from a linguistic standpoint.

Editors: Heidi Johnson (mailto:ailla@ailla.org)
Helen Aristar Dry (mailto:hdry@linguistlist.org)

Changes since previous version: Adds: transcription/phonemic, transcription/kinesic, annotation/translation, annotation/phonological, annotation/semantic, annotation/eye-gaze, annotation/facial-expression, description/phonological, description/kinesic, description/pedagogical, description/comparative, dataset/kinesic.

Deletes: transcription/eye-gaze, transcription/facial-expression, transcription/translation, transcription/phonological, transcription/semantic, description/eye-gaze, description/facial-expression, dataset/eye-gaze, dataset/facial-expression. Genre Type section.

Table of contents



Language Data Ontology

This is an experimental language data ontology based on OLAC terms for use in the ATAP and LDaCA projects

Classes

PrimaryText | Annotation | CollectionEvent | CollectionProtocol | DerivedResource | OrganizationBasedLicense | OrganizationDepositLicense | OrganizationReuseLicense | PersonSnapshot

Properties

annotationOf | annotationType | annotator | author | channels | collectionEventType | collectionProtocolType | compiler | consultant | dataInputter | depositor | derivationOf | developer | doi | editor | geoJSON | hasAnnotation | hasDerivation | illustrator | indexableText |
interpreter | interviewee | interviewe
recorder | register | researchParticip
hasCollectionProtocol | isDeIdentifi

DefinedTerms

Coded | Dialogue | Drama | Elicitatio
Orthographic | PartOfSpeech | Phon
Song | SpokenLanguage | Syntactic
WrittenLanguage | WhistledLangua

DefinedTermsSets

Class: PrimaryText

This is a primary resource: the object of study, such as a literary work, film, or recording of natural discourse

Property: linguisticGenre

A linguistic classification of the genre of this resource

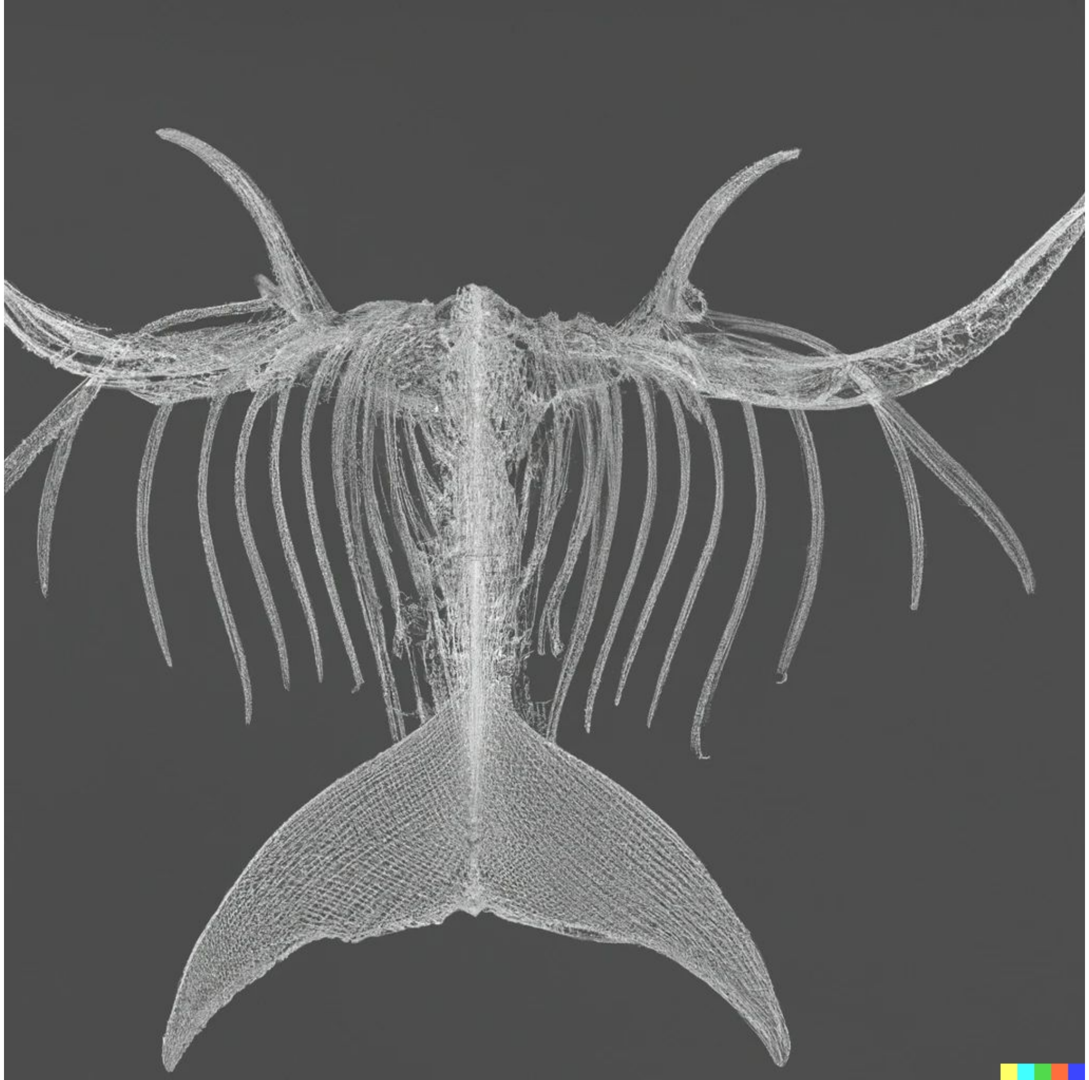
Values expected to be one of these types:

Used on these types:

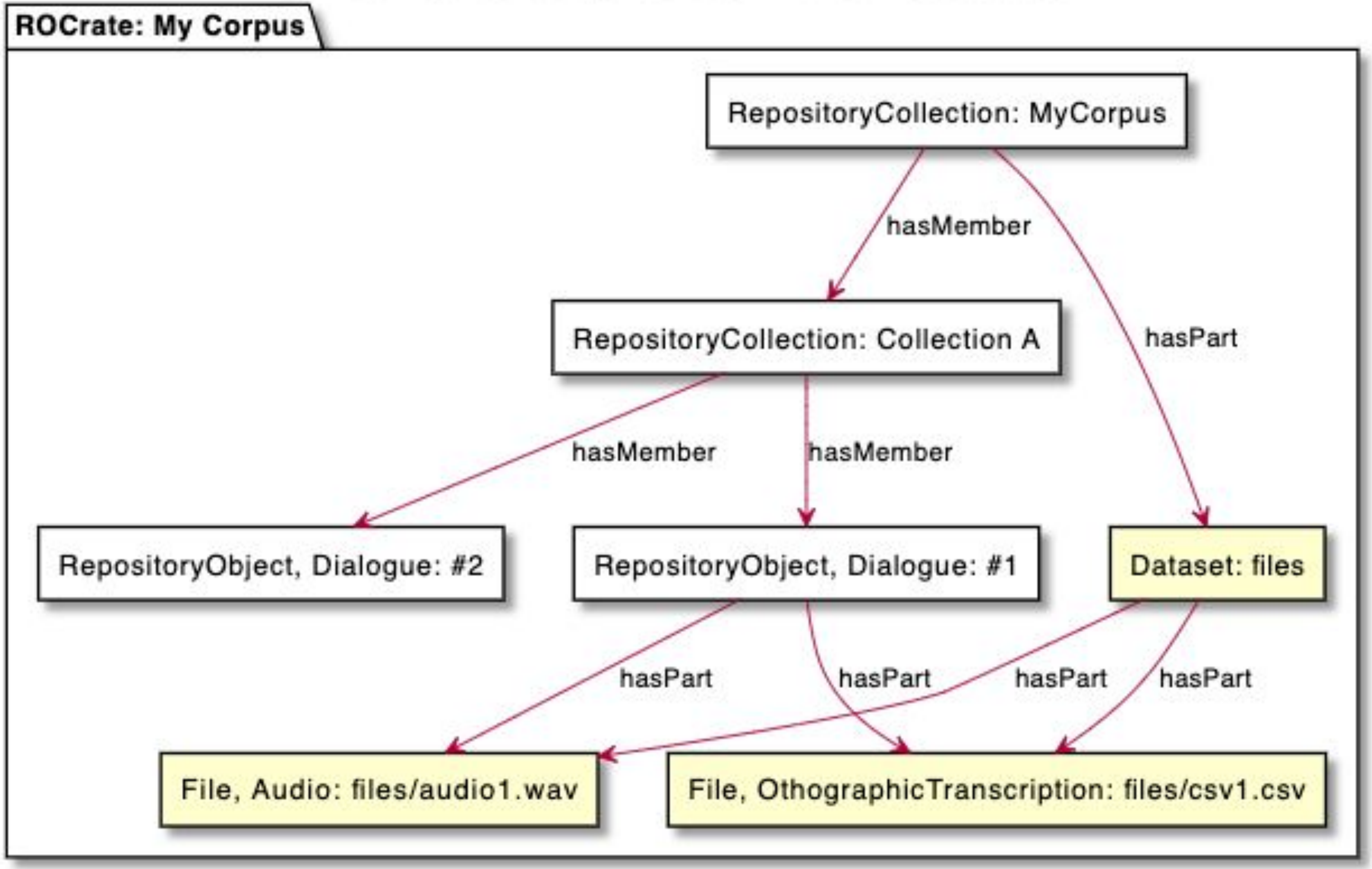
[https://purl.archive.org/language-data-commons/terms#PrimaryText] |

Values expected to be one of these defined terms:

[Formulaic] | [Thesaurus] | [Dialogue] | [Oratory] | [Report] | [Ludic] | [Procedural] | [Narrative] | [Interview] | [Drama] | [Informational] |



Self contained corpus crate with all resources



ROCrate: My Corpus

RepositoryCollection: MyCorpus

hasMember

hasPart

RepositoryCollection: Collection A

hasMember

hasMember

RepositoryObject, Dialogue: #2

RepositoryObject, Dialogue: #1

Dataset: files

hasPart

hasPart

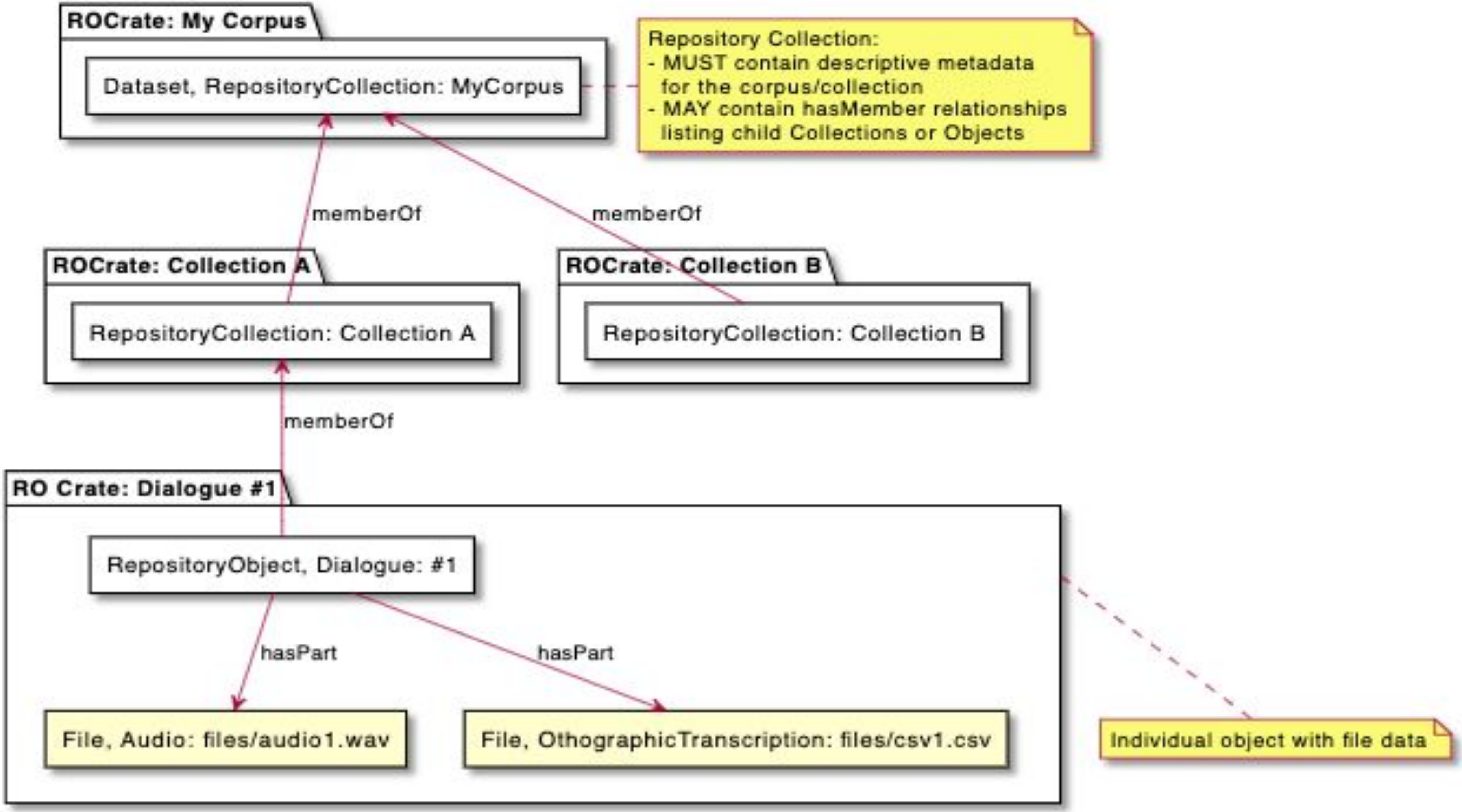
hasPart

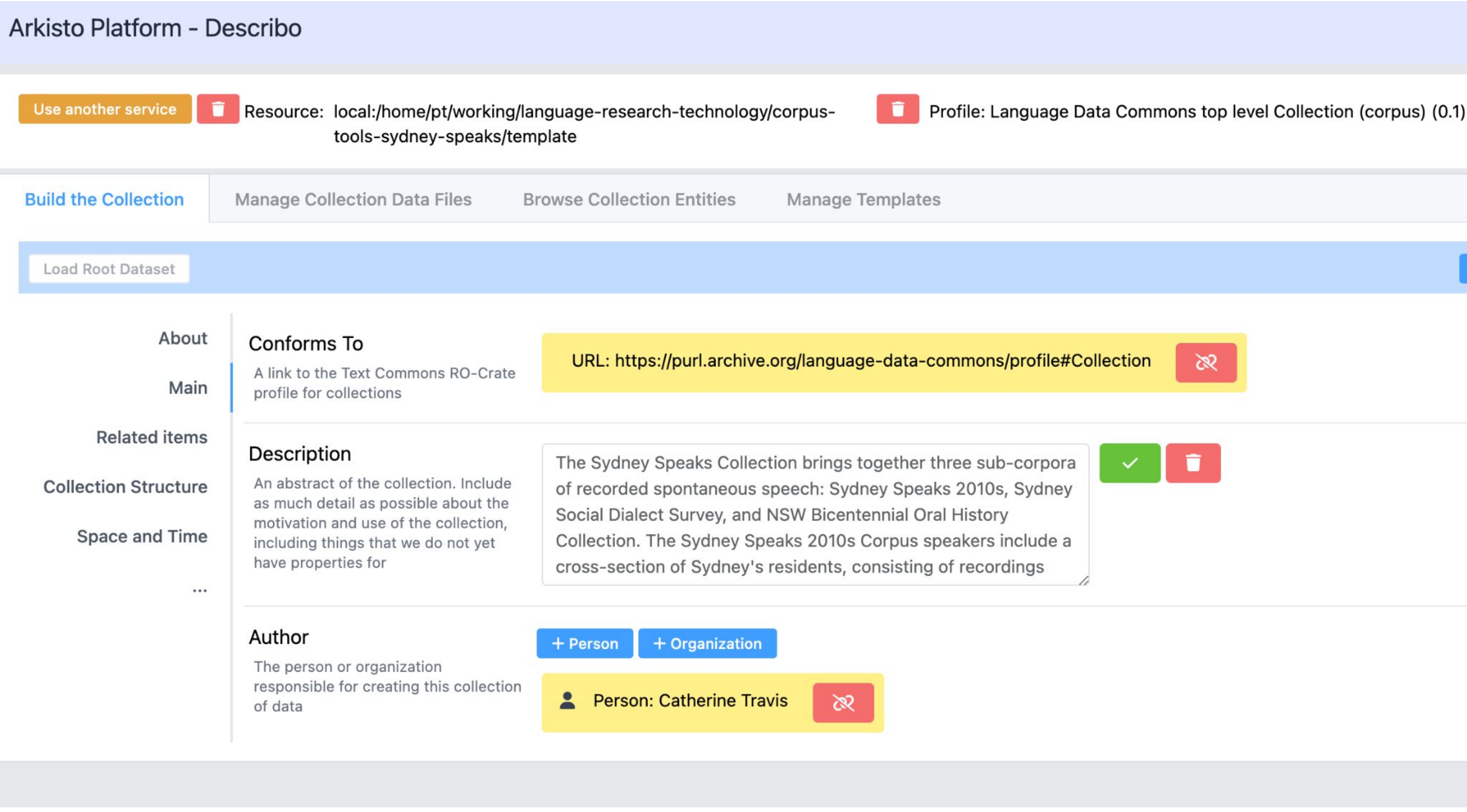
hasPart

File, Audio: files/audio1.wav

File, OthographicTranscription: files/csv1.csv

Complete corpus collection metadata-only crate w/ links to object packages


ROCrate: My Corpus
Dataset, RepositoryCollection: MyCorpus
Repository Collection:
- MUST contain descriptive metadata for the corpus/collection
- MAY contain hasMember relationships listing child Collections or Objects
memberOf
memberOf
ROCrate: Collection A
RepositoryCollection: Collection A
ROCrate: Collection B
RepositoryCollection: Collection B
memberOf
RO Crate: Dialogue #1
RepositoryObject, Dialogue: #1
hasPart
hasPart
File, Audio: files/audio1.wav
File, OthographicTranscription: files/csv1.csv
Individual object with file data



Resource: local:/home/pt/working/language-research-technology/corpus-tools-sydney-speaks/template

Profile: Language Data Commons top level Collection (corpus) (0.1)

Conforms To

A link to the Text Commons RO-Crate profile for collections

URL: https://purl.archive.org/language-data-commons/profile#Collection

Description

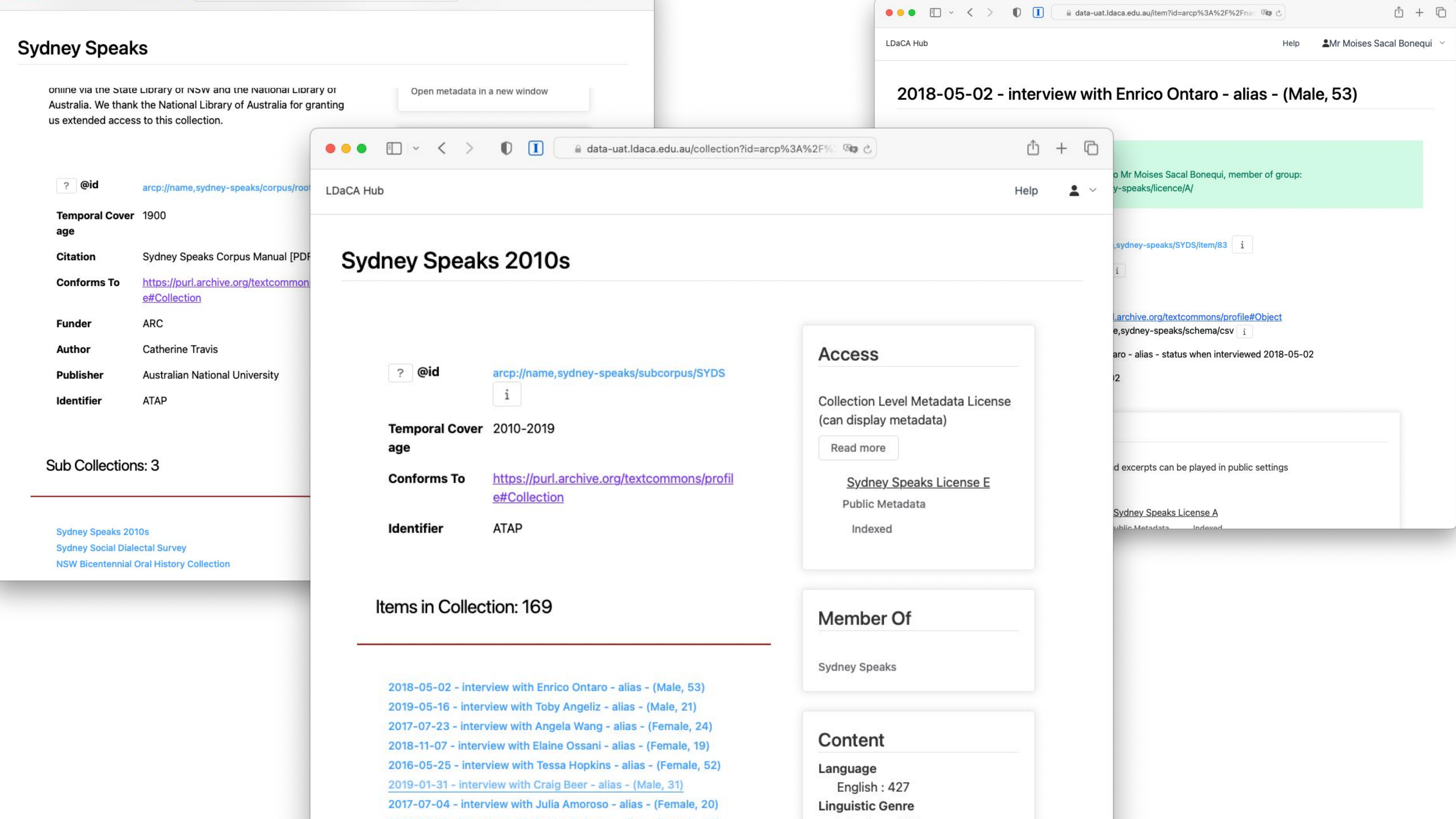
An abstract of the collection. Include as much detail as possible about the motivation and use of the collection, including things that we do not yet have properties for

The Sydney Speaks Collection brings together three sub-corpora of recorded spontaneous speech: Sydney Speaks 2010s, Sydney Social Dialect Survey, and NSW Bicentennial Oral History Collection. The Sydney Speaks 2010s Corpus speakers include a cross-section of Sydney's residents, consisting of recordings

Author

The person or organization responsible for creating this collection of data

Person: Catherine Travis



Sydney Speaks

online via the State Library of NSW and the National Library of Australia. We thank the National Library of Australia for granting us extended access to this collection.

? @id	arcp://name,sydney-speaks/corpus/root
Temporal Coverage	1900
Citation	Sydney Speaks Corpus Manual [PDF
Conforms To	https://purl.archive.org/textcommons e#Collection
Funder	ARC
Author	Catherine Travis
Publisher	Australian National University
Identifier	ATAP

Sub Collections: 3

- Sydney Speaks 2010s
- Sydney Social Dialectal Survey
- NSW Bicentennial Oral History Collection

Sydney Speaks 2010s

? @id	arcp://name,sydney-speaks/subcorpus/SYDS i
Temporal Coverage	2010-2019
Conforms To	https://purl.archive.org/textcommons/profile#Collection
Identifier	ATAP

Items in Collection: 169

- 2018-05-02 - interview with Enrico Ontaro - alias - (Male, 53)
- 2019-05-16 - interview with Toby Angeliz - alias - (Male, 21)
- 2017-07-23 - interview with Angela Wang - alias - (Female, 24)
- 2018-11-07 - interview with Elaine Ossani - alias - (Female, 19)
- 2016-05-25 - interview with Tessa Hopkins - alias - (Female, 52)
- 2019-01-31 - interview with Craig Beer - alias - (Male, 31)
- 2017-07-04 - interview with Julia Amoroso - alias - (Female, 20)

Access

Collection Level Metadata License (can display metadata)

Sydney Speaks License E

Public Metadata

Indexed

Member Of

Sydney Speaks

Content

Language

English : 427

Linguistic Genre

2018-05-02 - interview with Enrico Ontaro - alias - (Male, 53)

o Mr Moises Sacal Bonequi, member of group: y-speaks/licence/A/

,sydney-speaks/SYDS/item/83 i

i

.archive.org/textcommons/profile#Object

e,sydney-speaks/schema/csv i

aro - alias - status when interviewed 2018-05-02

2

d excerpts can be played in public settings

Sydney Speaks License A

ublic Metadata Indexed

Demo

Starting repository: Australian-Text-Analytics-Platform/farms-to-freeways/main

If a repository takes a long time to launch, it is usually because Binder needs to create the environment for the first time.

Loading Farms to Freeways from the API and ro-crate metadata file

The Language Data Commons of Australia (LDaCA) packages all their data collections in an ro-crate. There is a metadata file called ro-crate-metadata.json that comes with every data collection and this is how we can obtain metadata on this collection of research objects.

The metadata file is in the json format, and so we'll be learning how to read a json file in this notebook.