Describo and RO-Crate - the FAIR data research helpers

Dr Marco La Rosa[1], Dr Peter Sefton[2]

1.   *PARADISEC and The University of Melbourne*
2.   *University of Technology Sydney*

Problem

Research projects don't always involve work practices that describe the data generated so that it meets the FAIR principles.
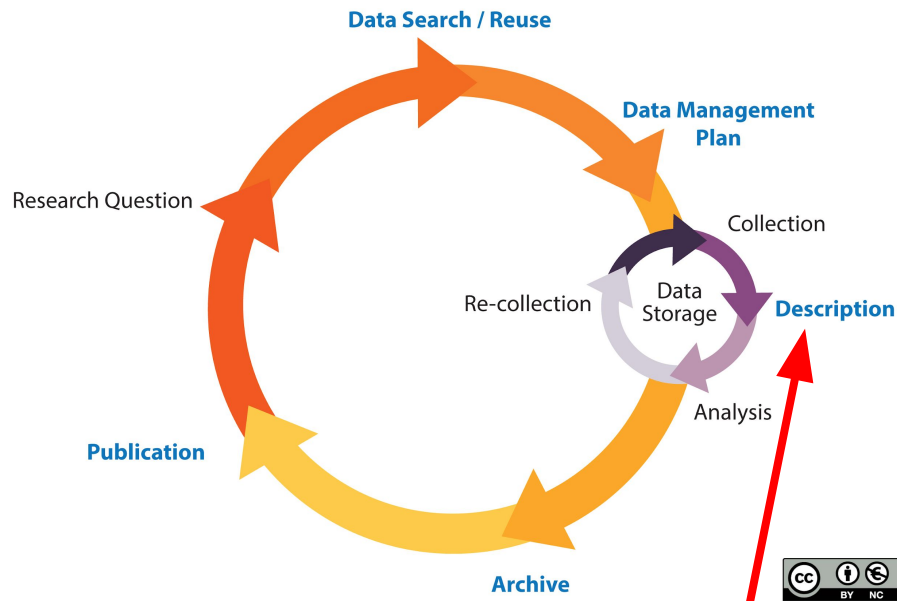
Solution

Describe the data as it's being created using open standards and tools

➔ use Describo

The Research Data Management Lifecycle

Data Search / Reuse

Data Management Plan

Research Question

Collection

Re-collection

Data Storage

Description

Analysis

Publication

Archive

Describo goes here

What is Research Object Crate (RO-Crate)?

- Lightweight approach to packaging research data with metadata

- Community effort - 38 contributors from AU, EU and US.

- Can aggregate files and/or any URI-addressable content, with contextual information to aid decisions about re-use. (Who What Where Why How).

- Uses Schema.org as the main ontology, with domain-specific extensions

➔ *Sefton is an Editor of the specification*

Who is using it?

- Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). Demonstrator with > 16k items and > 500 collections exported as RO-Crates into OCFL @ http://mod.paradisec.org.au/

- University of Technology Sydney

- Galaxy

- WorkflowHub

- CS3MESH4EOSC

Describo

- Describo is a pair of tools for researchers and support staff to create spec-compliant RO-Crates w/:
    a. Linked data for high-precision unambiguous metadata
    b. Package level "who, what, where" metadata
    c. Individual file-level metadata including provenance

- Describo exists as:
    a. a desktop tool (electron application) for all major platforms: https://uts-eresearch.github.io/describo/
    b. an online tool that integrates with Microsoft OneDrive: https://github.com/UTS-eResearch/describo-online

- Describo was developed as a proof of concept in two short sprints in 2020 (Desktop ~ 3 months and Online ~ 2 months). Both are viable products.

- https://arkisto-platform.github.io/assets/docs/Describo-WhitePaper.pdf

Describo Desktop - define and link entities

Author

🔗 + Person + Organization

👤 Person 🗑️
John Public

This item is connected to:

@type: RootDataset → hasPart 🗑️
name: my crate

Edit: Dataset - data

name ?
data

description ?
+ TextArea

license ?
🔗 + CreativeWork

hasPart ?
🔗 + Dataset + File

📁 Dataset 🗑️   📄 File 🗑️   📄 File 🗑️
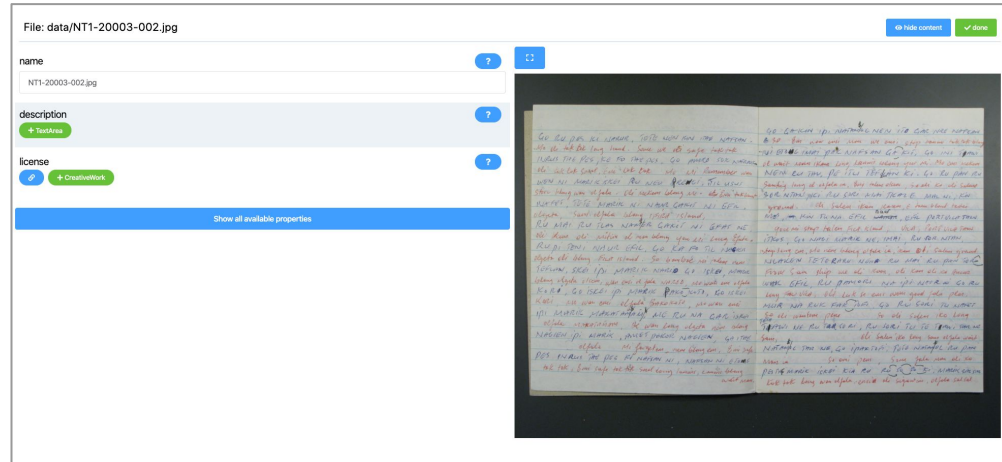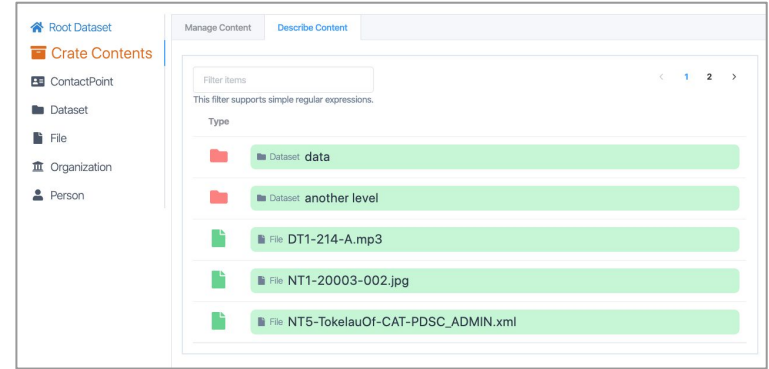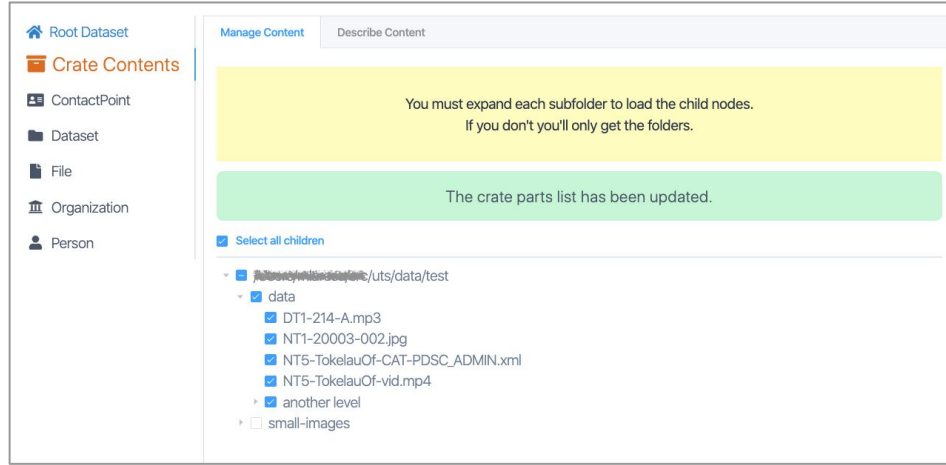another level   DT1-214-A.mp3   NT1-20003-002.jpg

📄 File 🗑️   📄 File 🗑️
NT5-TokelauOf-CAT-PDSC_ADMIN.xml   NT5-TokelauOf-vid.mp4

author ?
🔗 + Organization + Person

publisher ?
🔗 + Organization + Person

funder ?
🔗 + Organization + Person

Show all available properties
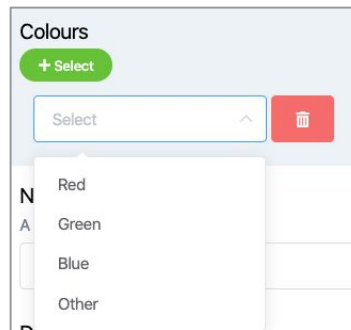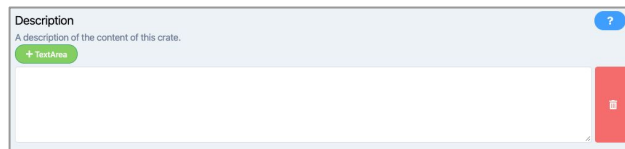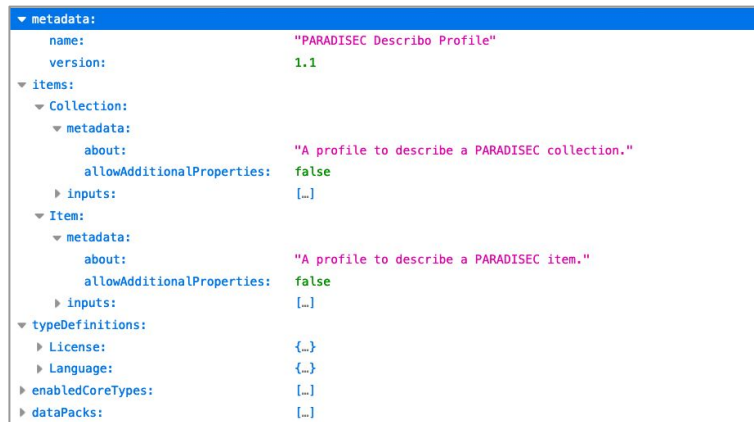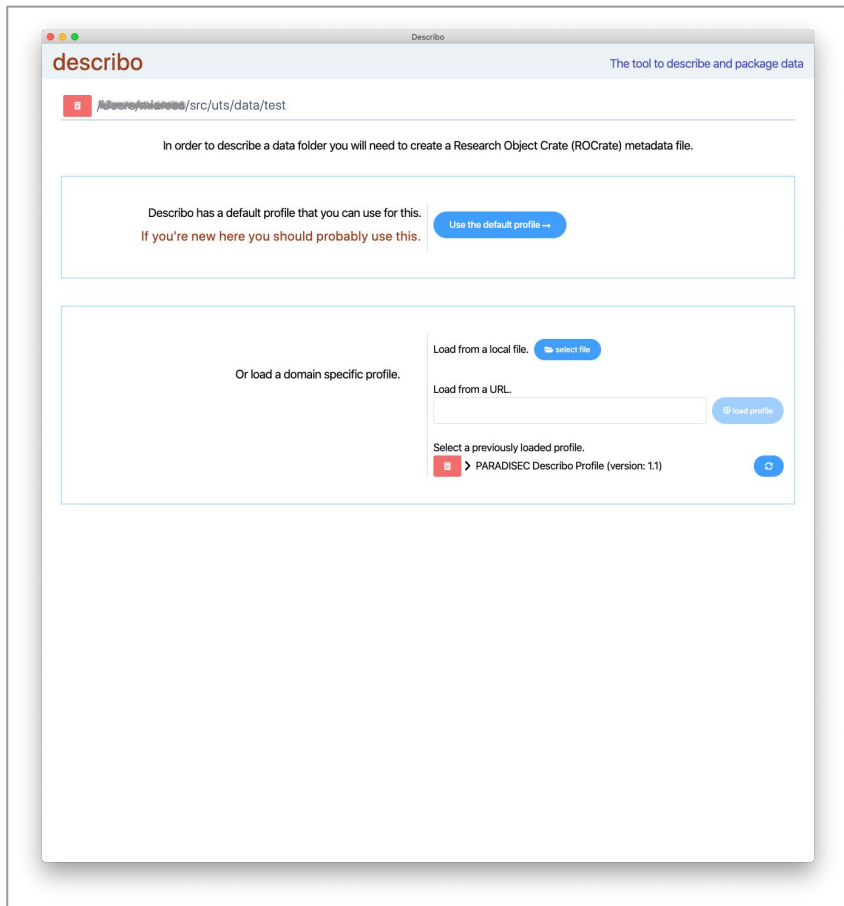
✓ save

This item is connected to:

@type: RootDataset → hasPart 🗑️
name: my crate

Describo Desktop - define and document data

Describo Desktop - load custom data profiles
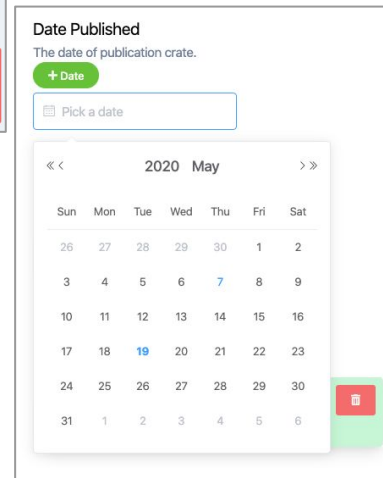
Describo Desktop - save / export

Crate

[inspect graph] [inspect root dataset] [inspect crate]

```
{
  "@context": "https://w3id.org/ro/crate/1.0/context",
  "@graph": [
    {
      "@id": "/ro-crate-metadata.jsonld",
      "@type": "CreativeWork",
      "about": {
        "@id": "./"
      },
      "identifier": "ro-crate-metadata.jsonld",
      "conformsTo": {
        "@id": "https://w3id.org/ro/crate/1.0"
      },
      "license": {
        "@id": "https://creativecommons.org/licenses/by-sa/3.0"
      }
    },
    {
      "@type": "Dataset",
      "name": "",
      "publisher": [
        {
          "@id": "https://ror.org/03f0f6041"
        }
      ],
      "contactPoint": {
        "@id": "#79082bdd-1db8-4c64-b141-1b6a80c170db"
      },
      "contentLanguages": [
        {
          "@id": "#3a9d3a13-cd7d-486d-82f4-ef0a95f437cf"
        }
      ],
      "hasPart": [
        {
          "@id": "data/"
        }
      ],
      "@id": "./"
    },
    {
```

Export my RO-Crate

A crate can be exported as a plain Zip archive or a BagIt Bag.

[📁 select folder]

Please select a location to save the export. You can't save the export to the same folder you're archiving or a subpath of it.

☑ **BagIt**
Should a BagIt bag be created?

test-2020-05-19-15-05.ro-crate
The name of the archive that will be created.

[Create Archive at /................/src/uts/data]

22%   [⊚ Stop Export]

Describo Online - architecture: integrate or standalone service

Easily deployed via docker containers: UI, API and PostgreSQL database. Start one instance and scale to many users or start small instances for individual domains.
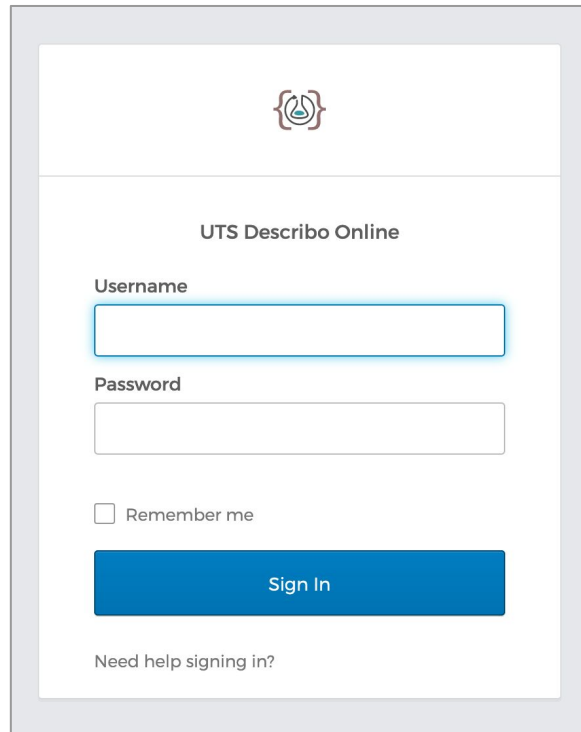
Architected for extensibility to other backends and data stores: S3, Dropbox, Google Drive, Openstack Swift, webdav, SFTP.

Operates on RO-Crate in database and synchronises back to target resource on change

Dynamically loads existing crates into DB and can associate crates on disk with in database entries to support multiple users operating on the data

User management and authentication
- ● via okta.com (developer.okta.com) (free up to 1000 Monthly Active Users)
- ● or integrate into your own application.

UTS Describo Online

Username

Password

☐ Remember me

Sign In

Need help signing in?

Describo Online - works with folders in Microsoft OneDrive
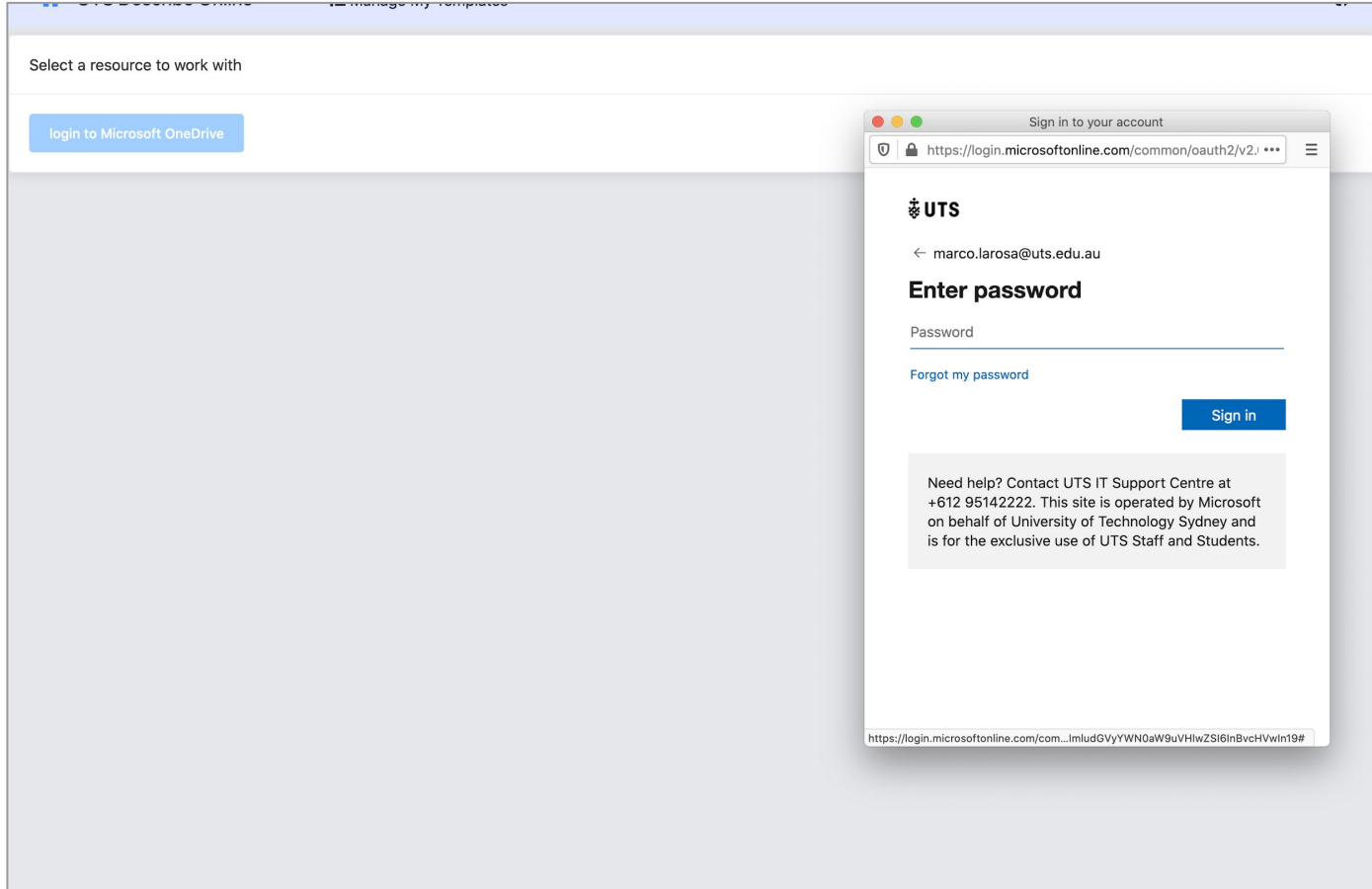
Describo Online - define and link entities

citation

Create and associate a new: **+ CreativeWork**

OR - Associate an existing 'CreativeWork': lookup by @id, @type, or name

ScholarlyArticle: This is an example publication with a dodgy DOI

Build the Collection   Manage Collection Data Files   Browse Collection Entities

Load Root Dataset   </> Add Property   Save Entity as Template   Delete Entity

@id   f261546b-cc25-46c0-8f1b-2eb4f8d5a210

@type   CreativeWork

name   new entity

This entity is referenced by:

📁 Dataset - Sample dataset for RO-Crate v1.0: citation

# Describo Online - manage crate entities

Describo Online - MVP - next steps

- Develop other backend integrations

- Develop a UI for novice users - current UI = expert RO-Crate user

- Develop a UI to manage cultural collections

- Port capability to upload domain specific profiles from desktop

Describo Desktop - next steps

- Port online entity rendering code from online - implementation cleaner than initial implementation in desktop

- Port novice user and cultural collections UI from online

- Port backend integrations from online

https://arkisto-platform.github.io/

Standards Based

Scaleable

Sustainable

arkisto
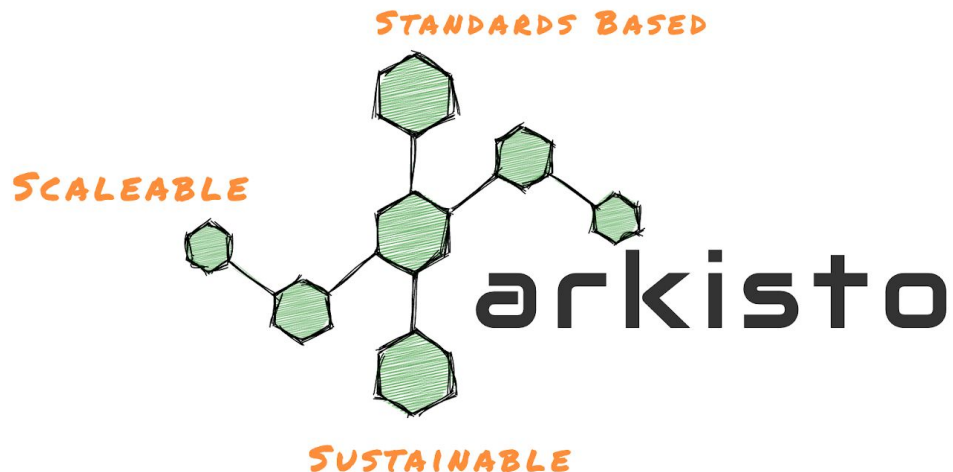
A scaleable, standards based platform for sustainable data.