Describo and RO-Crate - the FAIR data research helpers

Dr Marco La Rosa\textsuperscript{1}, Dr Peter Sefton\textsuperscript{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

Source: https://guides.library.ucsc.edu/datamanagement/
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 When Where Why How).
- Uses Schema.org as the main ontology, with domain-specific extensions
- Has human readable summaries of datasets

Sefton is an Editor of the specification: http://www.researchobject.org/ro-crate/
# Sample dataset for RO-Crate v1.1

<table>
<thead>
<tr>
<th>@id</th>
<th>/</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
<td>Sample dataset for RO-Crate v1.1</td>
</tr>
<tr>
<td>@type</td>
<td>Dataset</td>
</tr>
<tr>
<td>description</td>
<td>This is a simple dataset for demonstration purposes it contains just one image and a directory full of useless text files.</td>
</tr>
<tr>
<td>funder</td>
<td>Australian National Data Service University of Technology Sydney DataCrate Project</td>
</tr>
<tr>
<td>datePublished</td>
<td>2017-06-29</td>
</tr>
<tr>
<td>author</td>
<td>Peter Sefton</td>
</tr>
</tbody>
</table>
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 data repository

● Galaxy

● WorkflowHub: https://aboutworkflowhub.eu/

● 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/](https://uts-eresearch.github.io/describo/)
  b. an online tool that integrates with Microsoft OneDrive: [https://github.com/UTS-eResearch/describo-online](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 release candidates.

Describo Desktop - define and link entities

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

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

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

@id
@type
name

CreativeWork
new entity

This entity is referenced by:

Dataset - Sample dataset for RO-Crate v1.0: citation
Describo Online - manage crate entities

```
<table>
<thead>
<tr>
<th>@type</th>
<th>@id</th>
<th>Name</th>
<th>Connected</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>ContactPoint</td>
<td><a href="mailto:peter.sefton@uts.edu.au">mailto:peter.sefton@uts.edu.au</a></td>
<td>Contact Peter Sefton</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>CreateAction</td>
<td>#SepiaConversion</td>
<td>Converted dog picture to sepia</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>CreateAction</td>
<td>#Photo1</td>
<td>Took dog picture</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>CreativeWork</td>
<td><a href="https://creativecommons.org/licenses/by-nc-sa/3.0/au/">https://creativecommons.org/licenses/by-nc-sa/3.0/au/</a></td>
<td>Attribution-NonCommercial-ShareAlike 3.0 Australia (CC BY-NC-SA 3.0 AU)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>DataDownload</td>
<td>#picsdistro</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Dataset</td>
<td>pics/</td>
<td>Pictures</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
```
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/

A scaleable, standards based platform for sustainable data.