

# Describo

The FAIR data research helper

Dr Marco La Rosa

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

## The Research Data Management Lifecycle

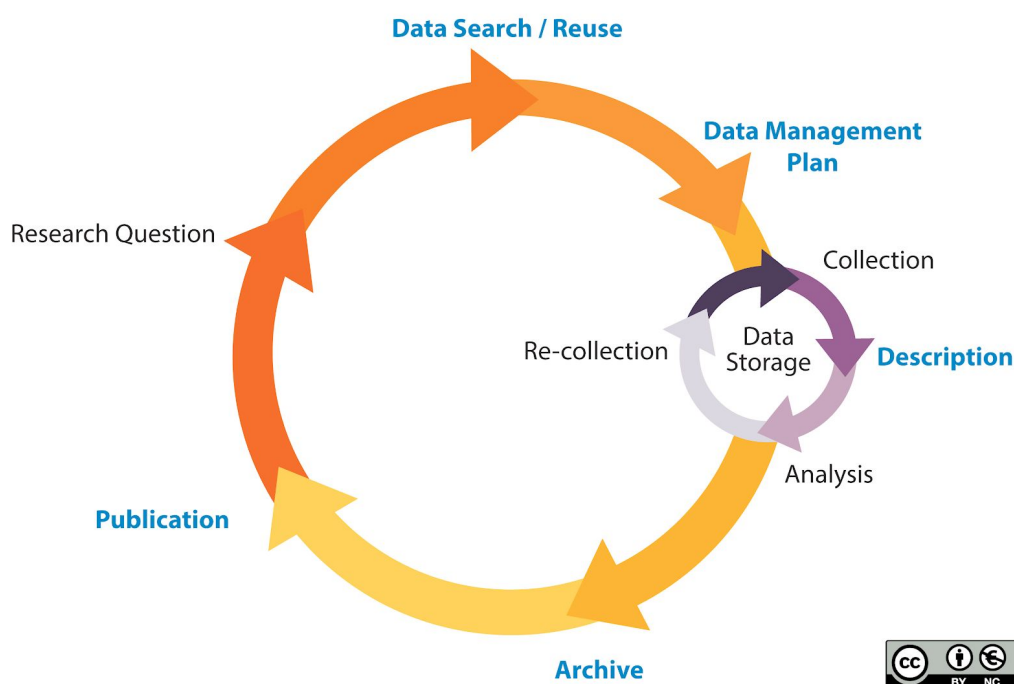


Figure 1. The Research Data Management Lifecycle.<sup>1</sup>

Research projects often don't involve work-practices that describe the data generated over the lifetime of the project so that it meets the FAIR principles: making data Findable so that it is Accessible, Reusable by others and Interoperable with archives, preservation systems and analytical systems.

This is especially true of smaller research projects where resource constraints may result in researchers using an application to compile and curate their data only to find out at the end of the project that the data and metadata are effectively locked up inside the application. Or worse, the research may be completed without any plan for the long term preservation of the data.

Further complicating this issue at an institutional level is how to preserve, and make accessible and discoverable, content coming from different sources and domains when the format of the incoming metadata varies by domain and ontology.

# Research Object Crate (RO-Crate)

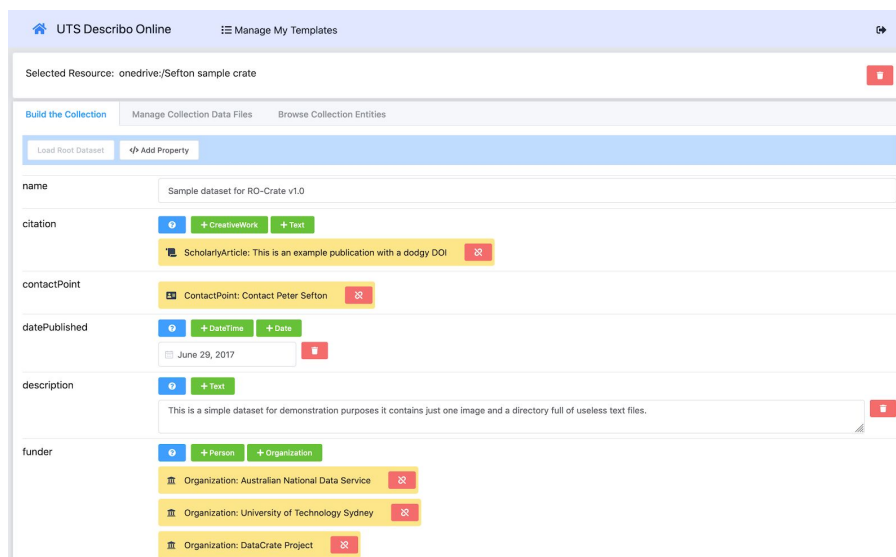
RO-Crate<sup>2</sup> is a way of packing data with machine-readable linked-metadata and a human-readable HTML file to make it FAIR compliant. RO-Crates can be indexed by standard web indexers meaning that the data becomes findable by Google Dataset search. Once found, it can be accessed and reused, informed by the metadata which is packaged alongside the data.

*RO-Crate is a community effort to establish a lightweight approach to packaging research data with their metadata. It is based on schema.org annotations in JSON-LD, and aims to make best-practice in formal metadata description accessible and practical for use in a wider variety of situations, from an individual researcher working with a folder of data, to large data-intensive computational research environments.”*

Source: <http://www.researchobject.org/ro-crate/>

Describo is an application that can be used by researchers and research support staff to create RO-Crate data packages.

## Describo



The screenshot shows the 'UTS Describo Online' interface. At the top, there's a navigation bar with 'Manage My Templates'. Below that, the 'Selected Resource' is 'onedrive/Sefton sample crate'. The main area is titled 'Build the Collection' and has tabs for 'Manage Collection Data Files' and 'Browse Collection Entities'. There are buttons for 'Load Root Dataset' and 'Add Property'. The form fields are:

- name:** Sample dataset for RO-Crate v1.0
- citation:** Includes a '+ CreativeWork' button and a '+ Text' button. A warning message says 'ScholarlyArticle: This is an example publication with a dodgy DOI'.
- contactPoint:** Includes a '+ ContactPoint' button and a '+ Text' button. A warning message says 'ContactPoint: Contact Peter Sefton'.
- datePublished:** Includes a '+ DateTime' button and a '+ Date' button. The date is set to 'June 29, 2017'.
- description:** Includes a '+ Text' button. The description is 'This is a simple dataset for demonstration purposes it contains just one image and a directory full of useless text files.'.
- funder:** Includes a '+ Person' button and a '+ Organization' button. Three organizations are listed: 'Organization: Australian National Data Service', 'Organization: University of Technology Sydney', and 'Organization: DataCrate Project'.

Figure 2. The Describo Online UI.

Describo is a tool to help researchers turn their folders of content into Research Object Crates suitable for sharing, reuse, and long term preservation in archival systems. Describo is open-source software.

Dr Peter Sefton developed the idea for Describo and is the product owner and primary client driver. The products are developed by Dr Marco La Rosa. Describo comes in two forms: a desktop application and an online version.

## Why Describo?

The data management lifecycle ideally looks something like that shown in Figure 1, in which “Description” could, and typically does, mean many things. Importantly for an institution, description as defined by different communities often results in data that is difficult to preserve and leverage at an institutional level.

And this is true only for those researchers / domains that have a data management plan and actively engage in data description. For many researchers / domains, resource constraints means Description is left as a project completion activity that may not actually occur.

Describo aims to slot into the “Description” section as a generic tool that implements the RO-Crate specification. Describo allows users to create crates for preservation, dissemination and re-use at the institutional level. That is, Describo directly supports the “Data Search / Reuse”, “Analysis”, “Archive”, and “Publication” stages of the research data management lifecycle depicted in Figure 1.

Indeed Describo supports a model where data is described and archived continuously rather than at the end of the project. And once described, the RO-Crate package can be shared with other researchers, put up on websites or uploaded to public repositories like figshare or zenodo.

## Describo Key Features

- Describe data in local folders (Describo Desktop) or Microsoft OneDrive (Describo Online). (Architected for extensibility to other backends and data stores: S3, Dropbox, Google Drive, Openstack Swift, SFTP).
- Describe the entities that define the context of your data: authors, contributors, organizations, places, etc.
- Describe contextual relationships.
- Preview content whilst describing it.
- Save templates (Describo Desktop and Online) for re-use in other crates.
- Export your crate as a zip file for sharing or preservation (Describo Desktop).
- Use all of schema.org (Describo Online) or use a domain specific profile (Describo Desktop).
- Integrate into institutional authentication via Okta or operate as part of another service.
- Modern UI / UX.
- Easy deployment and upgrades via Docker containers (Describo Online)

## Where do I get it?

- Describo Desktop for all major platforms: <https://uts-eresearch.github.io/describo/>
- Describo Online: <https://github.com/UTS-eResearch/describo-online>

## References

1. <https://guides.library.ucsc.edu/datamanagement/>
2. <http://www.researchobject.org/ro-crate/>