# Describo and RO-Crate - the FAIR data research helpers

Dr Marco La Rosa<sup>1</sup>, Dr Peter Sefton<sup>2</sup>

1. PARADISEC and The University of Melbourne

2. University of Technology Sydney



For research data to be **F**indable, and Interoperable and **R**eusable, it has to have metadata that makes it so - and this needs to be both human and machine readable so that finding aids can be built and people can search, or write programs to discover and use data. Also for research data to be distributed it has to be \*packaged\* into distributable objects - otherwise there is no digital object. This talk looks at one tool that implements the Research Object Crate standard that supports such FAIR packaging.

(NOTE That yesterday in the first session of the conference there were several interesting tools presented, including Rucio that gives access to data in Jupyter notebooks - but there was very little discussion of metadata, or data management systems, possibly because the large-scale physics data is already well managed? But we know that most research data is not, and there is a risk that if we provide share and sync services we will end up with each researcher having a petabyte of life-time's research data that's not at all organized.)

In an ideal data management lifecycle Describo slots into the Description section of the lifecycle.

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

RO-Crate brought together work by the Research Object team based in Manchester and work on data packaging in Sydney. It started as a means of packaging file-based data, . but as it uses JSON-LD as a serialization for linked-data metadata it can also aggregate any URI-addressable resource, or if necessary describe physical objects. RO-Crates have a metadata file describing a research object, and optionally its content and useful context such as who created the objects, using what equipment, software and workflows, and how to reuse the data and under which license(s). The idea is that if a user finds a zip file on their hard drive, and opens it there is an HTML file within which explains the data enough for them to be able to reuse it.

♠ Sample dataset for RC	-Crate v1.1
Sample datase	et for RO-Crate v1.1 etadata for Sample dataset for RO-Crate v1.1 in JSON-LD format
Download Crate v1.1 Sample dataset	this dataset: Sample dataset for RO- for RO-Crate v1.1
@id	/
name [?]	Sample dataset for RO-Crate v1.1
@type	Dataset
description [?]	This is a simple dataset for demonstration purposes it contains just one image and a directory full of useless text files.
funder [?]	Australian National Data Service University of Technology Sydney DataCrate Project
datePublished [?]	2017-06-29
author [?]	Peter Sefton

RO-Crate can (and should) have a human-readable summary of the metadata in the data package (or on the a website or repository).

## 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
- <u>Galaxy</u>
- WorkflowHub: https://about.workflowhub.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: <u>https://uts-eresearch.github.io/describo/</u>
  - b. an online tool that integrates with Microsoft OneDrive: <u>https://github.com/UTS-eResearch/describo-online</u>
- 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.
- https://arkisto-platform.github.io/assets/docs/Describo-WhitePaper.pdf

Describo exists as a desktop application and an online application. Both create spec compliant RO-Crates.

The desktop application is useful for describing filesystem based objects attached to the local machine including share drives and cloud drives like Dropbox.

The online application is designed for group / institutional level deployment and it supports describing content in Microsoft OneDrive. The Online version is multi user.

Describo Desktop - define and link er	ntities	
	Edit: Dataset - data	
Author	name	2
+ Person + Organization	asta description +totkes	7
Person	license	?
- Feison	hasPart	2
John Public	Dataset     another level     DT1-214-A.mp3     T1-20003-002.jpg	
	NT5-TokelauOf-CAT-PDSC_ADMIN.xml NT5-TokelauOf-vid.mp4	
This item is connected to:	author	•
	publisher	?
@type: RootDataset → hasPart name: my crate	funder funder funder funder funder	2
	Show all available properties	
		✓ save
	This item is connected to: @type: RootJutationt → hasPert ■ name: my crate	

The UI of describo enables the user to define and describe context entities like People, Places, Organizations and many others. Entities can be Interlinked.



Describo allow the user to describe the content of the folder - ie image, audio and video content. Content previews are available whilst describing individual items.

		version:	"PARADISEC Describo Profile" 1.1						
e oscribo	the	<pre>▼ items:</pre>							
	The rest of department of the period of the	✓ metadata:							
Advaro/midrated/src/uts/data/test		about:	"A profile to describe a PARADISEC collection	an."					
In order to deportine a data folder you will need to organize	ate a Decearch Object Crate (BOCrate) metariata file	a clowed clona coroper cles:	[_]						
in order to describe a data todel you will reed to crea	se a research object crate (rocrate) metadata ne.	Tten:	1						
		- metadata:							
Describo has a default profile that you can use for this.		about:	"A profile to describe a PARADISEC item."						
If you're new here you should probably use this.	Use the default profile →	allowAdditionalProperties:	false						
		▶ inputs:	[]						
		<pre>v typeDefinitions:</pre>							
		License:	(-)						
		F canguage.	(-/						
	Load from a local file.	> dataPacks:	[]						
Or load a domain specific profile.	Load from a URL								
	(B kand profile								
	Select a previously loaded profile.  PARADISEC Describe Profile (version: 1.1)	Description A description of the content of this crate. + Taukkes		Date Publis The date of pu + Date	hed ablication	crate.			_
				🗄 Pick a da	.te				
		Colours		« <	20:	20 M	lay		>>
		+ Selec		Sun Mor	n Tue	Wed	Thu	Fri S	at
				26 27	28	29	30	1	2
		Selec		3 4	Б	6	7	8	e
				10 11	12	13	14	15	.6
		N Red		17 18	19	20	21	22	22
				17 10	10	20	A1 .	74 4	0
		A Gree	in la						
		A Gree	in	24 25	26	27	28 3	29 3	0

Describo desktop supports loading custom profiles. This enables a community to define which entities / properties are relevant to them in order to guide the user when describing their data. The UI of describo is driven by the profile.



Describo desktop can save and export described folders as plain zip files or baglt bags for upload to institutional repositories/archives.

### 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	
Usernar	ne	
Passwo	rd	
Rem	ember me	
	Sign In	
Need he	lp signing in?	

Describo Online is targeted at the group / institutional deployment. It uses okta for authentication or it can be integrated into existing services by accepting delegated authentication.

Describo Online has been architected for extensibility to other back file sync and store applications such as OwnCloud and it supports multi user access by operating on the data in a database and then synchronising that back to the remote resource.



Describo Online currently works with Microsoft OneDrive. This was a use case of UTS but the backend is extensible to other cloud resources via rclone.

itation	0		×
	Create and assoc	ciate a new: + CreativeWork	
	OR - Associate an	n existing 'CreativeWork': lookup by @id, @type, or name	
	· O - La - La - La - La - Thi		
	ScholariyArticle: Thi	s 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
	Load Root Dataset 4	6 Add Property 1261546b-cc25-46c0-8f1b-2eb4f8d5a210	Save Entity as Template
	Load Root Dataset 4 @id @type	Add Property     [261546b-cc25-46c0-8f1b-2eb4f8d5a210     CreativeWork	🗟 Save Entity as Template 🖀 Delete Entity

Describo Online supports the full schema.org schema by default.

প UTS Descri	ibo Online 📰 Manage	My Templates									0
Selected Resource:	onedrive:/Sefton sample crate										Ĩ
Build the Collection	Manage Collection Data Files	Browse Collection Entities									
Add an entity to th	is collection:										
				Total 434 <	1 2	3	4	5	6	44	>
@type	@id		Name			Co	nect	ed	Actio	ns	
ContactPoint	mailto:peter.sefton@uts.edu.au		Contact Peter Sefton				~		ľ	Î	
CreateAction	#SepiaConversion		Converted dog picture to sepia				×		ľ	Î	
CreateAction	#Photo1		Took dog picture				×		ľ	Ĩ	
CreativeWork	https://creativecommons.org/lic	enses/by-nc-sa/3.0/au/	Attribution-NonCommercial-ShareAlike 3.0 Australia (CC	C BY-NC-SA 3.0	AU)		~		ľ	Ĩ	
DataDownload	#picsdistro						~		ľ	î	
DataDownload	https://data.research.uts.edu.au les/src/downloads/sample.zip	ı/examples/ro-crate/examp					~		ľ	Î	
Dataset	pics/		Pictures				~		ß		

Describo Online makes it easy to find entities and update them.

### 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

Describo Desktop and Online have some difference. Future work will aim to bring the two more inline and ensure that all capabilities are supported in both environments.



Find out more by visiting the Arkisto site where we also talk about repository patterns and use cases.