Arkisto: a repository based platform for managing all kinds of research data

Peter Sefton¹, Marco La Rosa², Michael Lynch¹

- 1. University of Technology Sydney
- The University of Melbourne

This presentation is

FAIR

driven

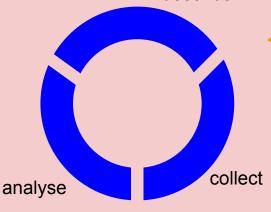
Research Data Management Plan

Workspaces:

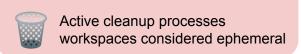
- working storage
- domain specific tools
- domain specific services

Reusable, Interoperable data objects

- → deposit early
- → deposit often



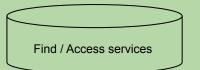
describe

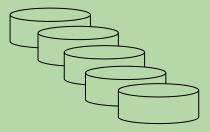




reuse data objects

Repositories: institutional, domain or both





Findable, Accessible, Reusable data objects



Policy based data management



Q. How can we "FIRST LOOK AFTER THE DATA"



Oxford Common File Layout Specification

Recommendation 07 July 2020

This version:

https://ocfl.io/1.0/spec/

Latest published version:

https://ocfl.io/latest/spec/

Editors:

Andrew Hanking Oniversity of Oxford)

Neil Jeffe , University of Oxford)

Rosal ersity)

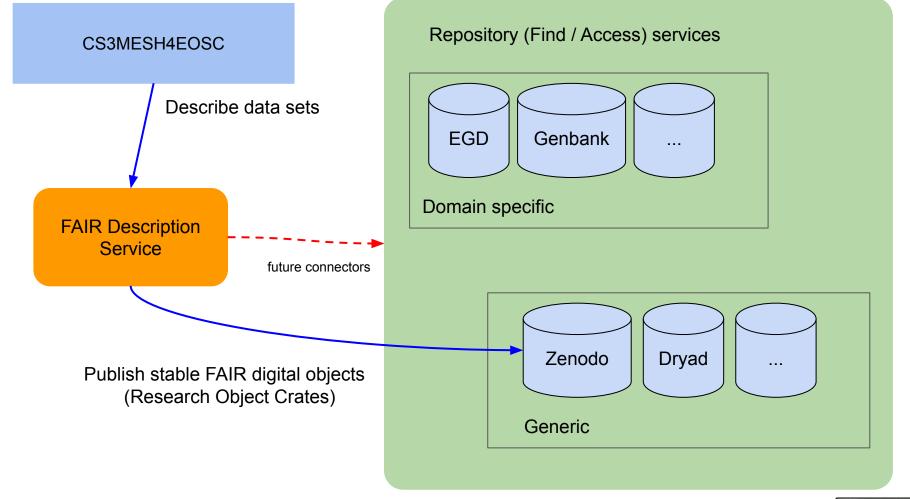
Julian I ord University)

Simeon \ er (Cornell University)

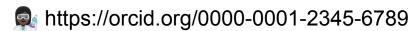
Andrew Woods (LYRASIS)



Additional Description





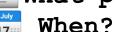




ID? Title? Description?



👰 晃 Who created this data? What parts does it have?





What is it about?



How can it be reused?



As part of which project?



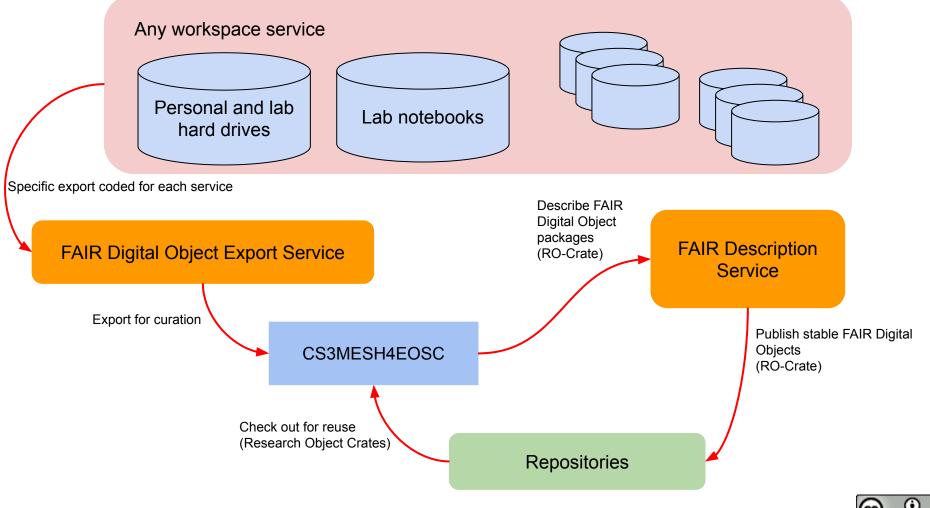
Who funded it?



How was it made?



https://en.wikipedia.org/wiki/Scanning electron microscope



Identity: authentication, authorisation and group services

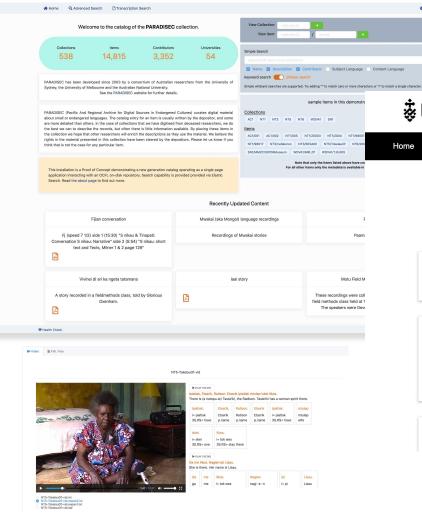
Provisioning

workspaces

repositories



UTS Provisioner Vocabulary Services ORCID Geo Server Data Portal 'Bot/s' for Automated Researcher Houskeeping Links to the Data Portal stash 3 Porta Research Information REDCap ≥ Elements Research Data Lookup Mgmt Plan Secure web application for REDCaP Adaptor Adaptor Research Data MINT Student Management building and managing online Catalogue technology (CASS) surveys and databases Adaptor Manage Congretation Labarchives 3 Provisioner A P115 Provisioner A P11 Visioner (Orchestrator) Other Future Research Apps Researcher cloud-based product to enable Storage Adaptor Future services as directed by Program researchers to store, organize, Adaptorls Ome GitLab Board eResearch Steering Committee and publish their research data Adaptor Adaptor MetaData Management & Provisionin® omeka High Performance Computing / Researcher Mid Range Computing Cross disciplinary and long-tail web based data repository amazon Researcher Microso Azure NCI Omero: Microscopy data management API cloudstor and virtual-lab services GitLab National Share / Sync collaboration linked to major facilities Researcher Code-centred research and data-analytics EMC' (ISILON' project management



Who is doing this?

UTS Library of Successful Grants

Climate Change Cluster



Reef Breath Testing (RBT): exhaled volatile-gas biomarkers of coral health

This Project aims to uncover volatile gas "fingerprints" of coral reef taxa and how they are diagnostic of healthy reef functioning over space and time. All organisms emit distinct volatile gases via physiological fine-tuning and signalling as their environments change. Whilst coral reef taxa and coral reefs are hotspots for volatile gas emissions, which gases are produced, when and why, is entirely unexplored. This project unites a multidisciplinary team of experts to, for the first time, couple volatile gas assessment, metabolic physiology and functional genomics techniques to transform understanding of how key volatile gases underpin coral resilience to stress and disease, which is essential to improve coral reef ecosystem management.

Search

Discovery Projects I 2020 I David John Suggett I Climate Change Cluster I 0306 - 03chemical Sciences;biological Sciences I 0602 - Ecology I 0699 - Other Biological Sciences I 0399 - Other Chemical Sciences I UTS Confidential

Diagnosing coral health tipping points under accelerating coastal hypoxia

Climate change and land use development are rapidly deoxygenating shallow water coral reefs and amplifying hypoxia exposure. This Project aims to unlock the role hypoxia plays in shaping the healthy functioning of corals over space and time.

Discovery Projects I 2018 I A/Prof David Suggett I Climate Change Cluster I 05 - Environmental Sciences I 06 - Biological Sciences I 0501 - Ecological Applications I 0602 - Ecology I 0608 - Zoology I UTS Confidential



OCFL Spec: https://ocfl.io/

Research Object Crate (RO-Crate) Spec: http://www.researchobject.org/ro-crate

UTS: https://github.com/UTS-eResearch/ro-crate-js

OCFL JS

- UTS OCFL JS Implementation: https://github.com/uts-eresearch/ocfl-js
- CoEDL OCFL JS implementation: https://github.com/CoEDL/ocfl-js

UTS RO Crate / SOLR portal: https://github.com/uts-eresearch/oni-express

Describo:

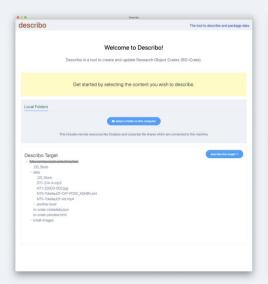
- https://github.com/Arkisto-Platform/describo
- https://github.com/Arkisto-Platform/describo-online
- https://github.com/Arkisto-Platform/describo-data-packs

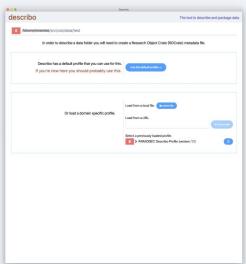
CoEDL Modern PARADISEC: https://github.com/CoEDL/modpdsc

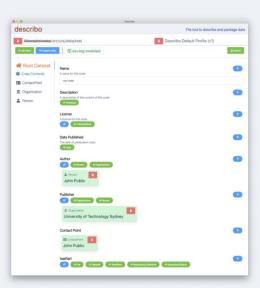
CoEDL OCFL tools: https://github.com/CoEDL/ocfl-tools

Describo The tool to describe and package data.

Describo is a tool to create and update Research Object Crates (RO-Crate)







Select a folder to describe.

Select a profile.

Describe your data.

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

arkisto-platform.github.jo



Why Arkisto

About

Standards

Storage

Packaging

Identifiers

Case Studies

PARADISEC

UTS Data

Grants

UTS Cultural Data

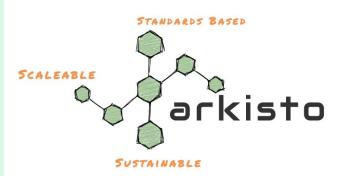
Use Cases

Tools

Data Description

Data Discovery

Data Import



A scaleable, standards based platform for sustainable data.

The basis of Arkisto is that the long-term preservability Microsoft Remote Desktop 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