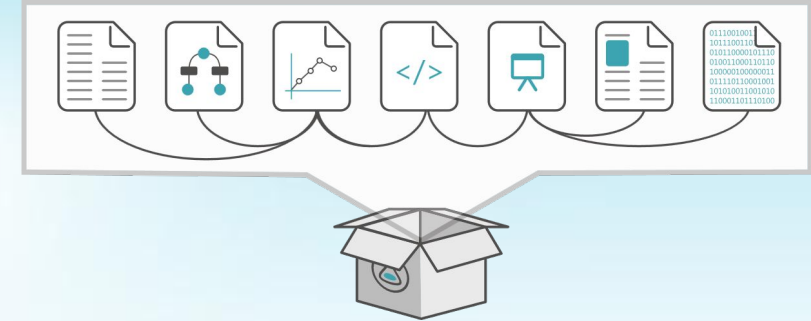




# RO-Crate and BioDT

Stian Soiland-Reyes, The University of Manchester

# Aims of FAIR Research Objects



**Describe** and **package** data collections, datasets, software etc.

→ data moves with its **metadata**

**Platform-independent** object exchange between repositories and services

Support **reproducibility** and **analysis**: link data with codes and workflows

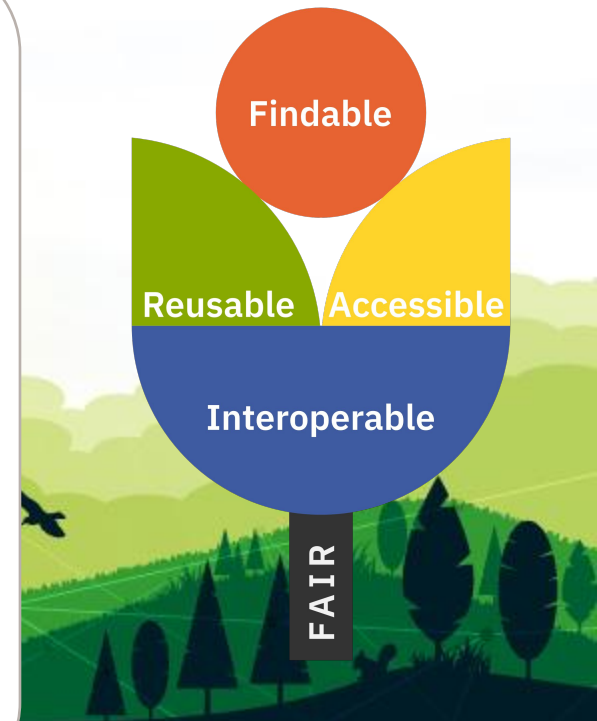
Transfer of **sensitive/large** distributed datasets with persistent identifiers

Aggregate **citations** and **persistent identifiers**

Propagate **provenance** and **existing metadata**

Publish and archive **mixed objects** and references

Reuse existing **standards**, but hide their complexity



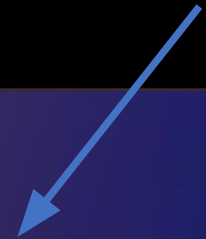


**BIODT**  
biodiversitydigitaltwin

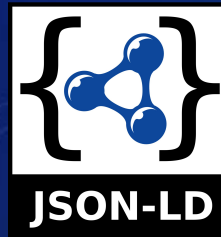
```
{
  "@id": "figure.png",
  "@type": ["File", "ImageObject"],
  "name": "XXL-CT-scan of an XXL Tyrannosaurus rex skull",
  "identifier": "https://doi.org/10.5281/zenodo.3479743",
  "author": {"@id": "https://orcid.org/0000-0002-8367-6908"},
  "encodingFormat": "image/png"
}
```



```
{
  "@id": "https://orcid.org/0000-0002-8367-6908",
  "@type": "Person",
  "affiliation": { "@id": "https://ror.org/03f0f6041" },
  "name": "J. Xuan"
}
```



```
{
  "@id": "https://ror.org/03f0f6041",
  "@type": "Organization",
  "name": "University of Technology Sydney",
  "url": "https://www.uts.edu.au/"
}
```



**Linked Data.. by stealth**

**schema.org**



## Species distribution models for Lycopodium

[Download all the metadata for Species distribution models for Lycopodium in JSON-LD format](#)

[Check this crate](#)

## Species distribution models for Lycopodium

@id	./
name [?]	Species distribution models for Lycopodium
@type	Dataset
description [?]	Crop Wild Relatives distribution modeling workflow using ModGP for Lycopodium. Lycopodium is not not actually a crop species, but serves as a good test genus due to the limited amount of species in this genus (and thus the short workflow execution runtime). Lycopodium is a genus of clubmosses within the family Lycopodiaceae. These perennial, evergreen plants are characterized by their small, needle-like or scale-like leaves arranged in spirals or whorls along the stems. Lycopodium species reproduce via spores rather than seeds and often form dense mats in forests, particularly in moist, shaded environments.
datePublished [?]	2024-10-12T16:02:25.893Z
author [?]	<a href="#">Claus Weiland</a>
conformsTo [?]	<ul style="list-style-type: none"><li><a href="https://w3id.org/ro/wfrun/process/0.1">https://w3id.org/ro/wfrun/process/0.1</a></li><li><a href="https://w3id.org/ro/wfrun/workflow/0.1">https://w3id.org/ro/wfrun/workflow/0.1</a></li><li><a href="https://w3id.org/workflowhub/workflow-ro-crate/1.0">https://w3id.org/workflowhub/workflow-ro-crate/1.0</a></li></ul>
dateCreated [?]	2024-10-12T16:02:27.337Z
dateModified [?]	2024-10-12T16:02:25.893Z
hasPart [?]	<ul style="list-style-type: none"><li><a href="#">Lycopodium_SDMDData.RData</a></li><li><a href="#">main.log</a></li><li><a href="#">main.log</a></li><li><a href="#">Lycopodium densum</a></li><li><a href="#">Lycopodium deuterodensum</a></li><li><a href="#">workflow.yaml</a></li></ul>
keywords [?]	<ul style="list-style-type: none"><li>Lycopodium</li><li>ModGP</li><li>SDM</li><li>test</li><li>CWR</li></ul>
license [?]	<a href="https://spdx.org/licenses/CC-BY-SA-4.0">https://spdx.org/licenses/CC-BY-SA-4.0</a>

FAIR is not just  
machine-readable!

<https://www.npmjs.com/package/ro-crate-html>

# RO-Crate in practice

RO-Crate is used by multiple international projects

Applied across research domains – from **life sciences** to **cultural heritage**

[https://www.researchobject.org/ro-crate/use\\_cases](https://www.researchobject.org/ro-crate/use_cases)

## Arkisto



Arkisto was a project website that aimed to outline a principled approach to research data management.

## Autosubmit



Autosubmit is an open source Python experiment and workflow manager used to manage complex workflows on Cloud and HPC platforms. Autosubmit uses RO-Crate to package the configuration, traces (logs, metrics, databases, etc.), and data of experiments and workflows.

## BioConnect



JAX BioConnect is an index of research data that supports data sharing, high-quality curation, and consistent data description.

## COMPSS



The COMPSS programming model is able to record Workflow Provenance in RO-Crate format, for governance and reproducibility of computational experiments

## Data Plant

DataPLANT's mission is to provide a sustainable and well annotated data management platform for plant sciences.

## Data Stewardship Wizard



The Data Stewardship Wizard (DSW) is an interactive platform for making data management plans

## Dataverse & AROMA

AROMA (ARP RO-Crate Manager) is part of Hungarian initiative ELKH ARP, extending Harvard Dataverse to allow dynamic metadata editing of data deposit metadata.

## DeSci Nodes

The DeSci Nodes system has been developed by the DeSci foundation, where dPID (distributed Persistent Identifier) act as an overlay of the Interplanetary File System (IPFS)

## EGI AppDB

EGI's Application Database (AppDB) has support for RO-Crate download, either as a JSON-LD metadata file or a ZIP archive.

## FAIRSCAPE

FAIRSCAPE is a framework for reusable cloud-based computations using ARK identifiers with rich provenance in an evidence graph and the Evidence Graph Ontology (EVI)

## Five Safes Crate

The Five Safes RO-Crate profile extend the Workflow Run RO-Crate profile for use in Trusted Research Environments (TRE).

## KEDO Data Lake

Knowledge Enhanced Digital Objects (KEDO) is an experimental approach of building a data lake using a combination of knowledge graphs, RO-Crate and PID records.

## Language Data Commons of Australia



LdCA uses RO-Crate as an interchange and archive format for language data, and is providing data discovery portals and API access to data using RO-Crate-centric APIs.

## Life Monitor



Life Monitor is a testing and monitoring service for computational workflows being developed as part of the EOSC-Life project. It aims to facilitate the execution, monitoring and sharing of workflow tests over time, allowing to detect deviations from expected workflow operation and provide useful feedback to the workflow authors

## LivePublication

LivePublication is a proof of concept of an executable paper, which interactive visualization and statistical calculations can be regenerated on the fly taking into consideration data sources updated after the paper's publication date.

## Machine-actionable data management plans

RDA maDMP Mapper and Ro-Crate\_2ma-DMP can convert between machine-actionable data management plans (maDMP) and RO-Crate.

## PARADISEC

Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC) holds 16,100 hours of audio recordings and 2,800 hours of video recordings that might otherwise have been lost. These recordings are of performance, narrative, singing, and other oral tradition. This amounts to over 220 terabytes, and represents 1,370 languages, mainly from the Pacific region.

## PILARS

PILARS is a set of Protocols for Implementing Long-term Archival Repository Services

## ROHub



ROHub is a solution for the storage, lifecycle management and preservation of scientific work and operational processes via research objects. It makes these resources available to others, allows to publish and release them through a DOI, and allows to discover and reuse pre-existing scientific knowledge.

## RRkive



RRkive is a website aimed to outline a principled approach to research data management with guidance on data storage and metadata

## Research Object Composer

Research Object Composer is a REST API for gradually building and depositing Research Objects according to a pre-defined profile.

## Sciebo RDS

Sciebo RDS (Research Data Services) is a self-hosted interface between data repositories and file storage solutions, assisting the research data deposition process with annotations made using Describo Online and stored as an RO-Crate

## Survey Ontology

The Survey Ontology is an open vocabulary that allows representing, annotating and sharing a representation of the survey structure and the gathered responses.

## Time Layered Cultural Map (TLCMap)



TLCMap is a set of tools that work together for mapping Australian history and culture which includes downloads of geographical data packaged in RO-Crate

## UTS Cultural Datasets

The UTS Cultural Datasets project is collaborating with Humanities and Social Science (HASS) researchers and is re-using existing UTS Data infrastructure to build interactive services that allow people to use the data.

## UTS Research Data Repository

The UTS Research Data Repository is a searchable portal for discovering and accessing public datasets by UTS researchers.

## Wildlive portal

Senckenberg's Wildlive data portal, is a repository and analysis platform for biodiversity monitoring.

## Workflow Execution Service (WfExS)



WfExS-backend is a high-level workflow execution command line program that consumes and creates RO-Crates, focusing on the interconnection of content-sensitive research infrastructures for handling sensitive human data analysis scenarios

## WorkflowHub



WorkflowHub imports and exports Workflow RO-Crates, using it as an exchange format. They are a specialization of RO-Crate for packaging an executable workflow with all necessary documentation. It is aligned with, and intends to strictly extend, the more general Bioschemas ComputationalWorkflow profile.

## ZBMed SemTec web pages



The Semantic Technologies (SemTec) team in ZB MED uses GitHub pages to share research projects and corresponding research artefacts/outcomes.



# Collecting corpora for a Language Data Commons



THE UNIVERSITY OF QUEENSLAND AUSTRALIA

https://language-research-technology.github.io/crate-o/#?id=dx.doi.org%2F10.25911%252F42ds-nh06

Crate-O v0.3.10

Open Directory Load Files Bulk Add Save Close Help About

Mode: Language Data Commons top-level Collection (corpus) Selected Directory: try01

Corpus of Australian...nd Language (AusESL)

Current Entity: **Corpus of Australian English as a Second Language (AusESL)**

About

Related People, Orgs & Works

Structure

Provenance

Space & Time

Software & Hardware

Others

Conforms to

Name

Description

dx.doi.org/10.25911/42ds-nh06

Persistent, managed unique ID in URL format (if available), for example a DOI for a collection or an ORCID, personal home page URL or email address for a person

@type Dataset RepositoryCollection

RepositoryCollection

Corpus of Australian English as a Second Language (AusESL)

Organization/Business/Institution Accented Australian English License

File Annotation Text\_Anonymised(AusESL\_MF\_011\_Rosa.txt)

RepositoryObject AusESL\_MM\_044\_Xiaobo

RepositoryObject AusESL\_MF\_024\_Liu

RepositoryObject AusESL\_MF\_043\_Lanfen

RepositoryObject AusESL\_RF\_023\_Valerie

RepositoryObject AusESL\_RF\_007\_Melania

Enter keyword to filter the values

1 2 3 4 ... 54 Total 539

RepositoryObject AusESL\_MF\_011\_Rosa

Person Rosa

Dataset RepositoryCollection

Corpus of Australian English as a Second Language (AusESL)

Organization/Business/Institution Accented Australian English License

File Annotation

Text\_Anonymised(AusESL\_MF\_011\_Rosa.txt)

RepositoryObject AusESL\_MM\_044\_Xiaobo

RepositoryObject AusESL\_MF\_024\_Liu

RepositoryObject AusESL\_MF\_043\_Lanfen

RepositoryObject AusESL\_RF\_023\_Valerie

RepositoryObject AusESL\_RF\_007\_Melania

The type of the entity.

A link to the Text Commons RO-Crate profile for collections.

The name of this dataset.

The corpus of Australian English as a Second Language (AusESL) comprises audio-recorded sociolinguistic interviews with people who migrated to Australia as adults from non-English-speaking countries and have been living in Australia for at least two years. As of January 2024, the corpus has recordings with 30 Russian speakers (19 women & 11 men) and 30 Mandarin speakers (20 women & 10 men). The two groups are comparable in terms of age (average 36 years), age of arrival in Australia (average 26 years), and length of residence (average 11 years). All are tertiary-educated, and over two thirds have a post-graduate degree (N = 42). All speak English fluently, though with varying degrees of proficiency. The interviews were conducted in English by

An abstract of the collection. Include as much detail as possible about the motivation and use of the dataset, including things that we do not yet have properties for.

```
{
  "@id": "#place-Kiriwina",
  "@type": "Place",
  "geo": {
    "@id": "#geo-151.0,-8.6-151.2,-8.6-151.2",
    "name": "Kiriwina"
  },
  "@id": "#language-kij",
  "@type": "Language",
  "code": "kij",
  "location": {
    "@id": "#geo-150.294,-8.879-151.2,-8.879-151.2",
    "name": "Kilivila"
  },
  "@id": "#geo-150.294,-8.879-151.2,-8.879-151.2",
  "@type": "Geometry",
  "asWKT": "POLYGON((150.294 -8.32, 151.2 -8.879, 150.294 -8.879, 150.294 -8.32))"
}
```

https://data.ldca.edu.au

Home Ldca

Collections Notebooks Browse Dr Peter Sefton Help

Search...

Advanced Search beta

RESET SEARCH Sort by: Collections Order by: Descending

1 2 3 4 5 6 ... 3879

Filters

Main Collections

Filter

The speech of Australian adolescents: research data and recordings collected by A.G. Mitchell and Arthur Delbridge in 1959 and 1960 (5940)

A COpus of Oz Early English (COOEE) (4072)

Australian Corpus of English (ACE) (1268)

International Corpus of English (ICE-AUS) (1129)

Braided Channels (60)

Sub-Collection

Access

Record Type

Language

Communication Mode

Linguistic Genre

Total: 38786 Index entries (Collections, Objects, Files and Notebooks)

Australian Deafblind Signing Corpus

Type: Dataset RepositoryCollection

Language: English

Australian Deafblind Signing Corpus captures free conversations between consenting participants as well as discussions at two workshops. The vast majority of the free conversations in the corpus are between two deafblind people, but we also recorded conversations between our focal participant...

See more

The La Trobe Corpus of Spoken Australian English

Type: Dataset RepositoryCollection

Language: English

The La Trobe Corpus of Spoken Australian English (LCSAusE) comprises a collection of six recordings and transcriptions of spoken interaction amongst Australian speakers of English (some in conversation with native French speakers speaking English) made in Melbourne from 2001 to 2002. The...

Objects: 12, Files: 24

See more

International Corpus of English (ICE-AUS)

Type: Dataset RepositoryCollection

Language: English

The Australian component of the International Corpus of English (ICE-AUS) is an approximately one million word corpus of transcribed spoken and written Australian English from 1992-1996. It consists of 500 samples of Australian English (60% speech, 40% writing) that matches the structure of other...

Collections: 12, Objects: 558, Files: 558

See more

A COpus of Oz Early English (COOEE)

Type: Dataset RepositoryCollection

Language: English

Material to be included had to meet with a regional and a temporal criterion. The latter required texts to have been produced between 1788 and 1900 in order to become eligible for COOEE. It was mandatory for a text to have been written in Australia, New Zealand or Norfolk Island. But in a few cases, other...

Objects: 1357, Files: 2714

<https://youtu.be/p-GZbe-Kzww>



<https://www.researchobject.org/ro-crate/ldca>

<https://www.researchobject.org/ro-crate/paradisec>

Browser tabs: OBM - Agenda, Galaxy | Europe, conda-forge | commu..., GCC CoFest 2024 - Cro..., Galaxy | Configured b..., Galaxy | Datarmor Test, Boîte de réception (3..., Home - Galaxy Comm..., Element [3] | galaxy-iu...

Address bar: <https://earth-system.usegalaxy.eu>

Galaxy Earth System | Workflow Visualize Data Help User




# Welcome to Galaxy for Earth System and environment

Galaxy for Earth System and environment was implemented within the project [Fair-Ease](#). It's a virtual platform to process, analyse and visualize Earth System, Environment and Biodiversity data. It is based on the [Galaxy framework](#), which guarantees simple access, easy extension, flexible adaption to personal and security needs, and sophisticated analyses independent of command-line knowledge.

Content <https://earth-system.usegalaxy.eu/>

History

search datasets

Marine Omics: SMBGC annotation

2.01 MB

- 26: 48574534
- 25: Sanntis output data
- 24: InterProScan on data 23 (ts v)
- 23: Regex Find And Replace on data 19
- 22: Sanntis output data genbank
- 21: Prodigal Gene Predictor on data 15 : complete starts file
- 20: Prodigal Gene Predictor on data 15 : nucleotide sequences file
- 19: Prodigal Gene Predictor on data 15 : protein translations file
- 18: Prodigal Gene Predictor on data 15 : coordinates
- 17: Sanntis output data
- 16: InterProScan on data 15 (ts v)



- Run the workflow



- Galaxy saves all the analytical steps of the workflow in your history


Invoked Workflow: "Marine Omics identifying biosynthetic gene cl..." ← Invocations List

*invoked 8 minutes ago*  
History: Marine Omics: SMBGC annotation

Workflow Version: 6 Edit  
workflow runs: 1 Run

Overview Inputs Outputs Report Export

Generate PDF 6 of 6 steps successfully scheduled.  
5 of 5 jobs complete.

**EOSC FAIR-EASE**  
Galaxy Earth system   
<https://earth-system.usegalaxy.eu/>  
Come try our earth system **subdomain** and test out


1: Fasta nucleotide file  
This is an input

- Step 1: Fasta nucleotide file
- Step 2: Prodigal Gene Predictor
- Step 3: Sanntis: Build Genbank
- Step 4: Regex Find And Replace
- Step 5: InterProScan
- Step 6: Sanntis: identify biosynthetic gene clusters

Overview Inputs Outputs Report **Export**

Last updated: 3 months ago ; Invocation ID: 96ca82ae449b2d13

## Research Object Crate (RO-Crate)

Generate  

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.

Learn more about RO Crate.

- 9: Retrieve ENA data on SAMEA 2620084
- 8: Retrieve ENA data on SAMEA 2620084
- 7: Retrieve ENA data on ERA00 0010
- 6: tac on data 1



Several workflow engines have implemented the [WRROC profiles](#) or are planning to do so:

Implementation	Profile	Version URL/DOI	Example
<a href="#">runcrate</a>	Provenance	<a href="#">runcrate 0.5.0 or later</a>	<a href="#">10.5281/zenodo.7774351</a>
<a href="#">Galaxy</a>	Workflow	<a href="#">Galaxy 23.1.1 or later</a>	<a href="#">10.5281/zenodo.7785861</a>
<a href="#">COMPSs</a>	Workflow	<a href="#">comps 3.2 or later</a>	<a href="#">10.5281/zenodo.7788030</a>
<a href="#">StreamFlow</a>	Provenance	<a href="#">Streamflow 0.2.0.dev10</a>	<a href="#">10.5281/zenodo.7911906</a>
<a href="#">WfExS</a>	Workflow	<a href="#">WfExS 0.10.1 or later</a>	<a href="#">10.5281/zenodo.10091550</a>
<a href="#">Sapporo</a>	Workflow	<a href="#">sapporo-service 1.5.1 or later</a>	<a href="#">10.5281/zenodo.10134581</a>
<a href="#">Autosubmit</a>	Workflow	<a href="#">Autosubmit v4.0.100 or later</a>	<a href="#">10.5281/zenodo.8144612</a>
<a href="#">Nextflow</a>	Provenance	<a href="#">(nf-prov in development)</a>	<a href="#">example</a>

## Workflow systems in BioDT

**Nextflow:** Assisting developers for RO-Crate of runs and from nf-core

**Snakemake:** With BGE adding RO-Crate of runs

**pydoit:** Created <https://github.com/BioDT/pydidit>

**Lexis:** Exploring two strands:

- a) Creating RO-Crate as part of workflow
- b) Airflow engine to save provenance of workflow execution

“Manual” workflows: Choose a template, use [Crate-O](#) editor, and upload to WorkflowHub

**RO-Crate native:** DestinE Platform – evaluate Galaxy, WfExS?



<https://workflowhub.eu/>

A registry for describing, sharing and publishing **scientific computational workflows**

WorkflowHub aims to **facilitate discovery and re-use** of workflows in an accessible and interoperable way. This is achieved through extensive use of **open standards** and tools, including [CWL](#), [RO-Crate](#), [Bioschemas](#) and [GA4GH's TRS API](#), in accordance with the **FAIR principles**.

WorkflowHub **supports workflows of any type** in its native repository.

[Learn more](#)

[Register](#)

Welcome to WorkflowHub

- **Help** is available on [about.workflowhub.eu](https://about.workflowhub.eu).
- Report any **issues** or **suggest new features** on [GitHub](#).
- For **comments, questions or feedback**, please use the [feedback form](#).

Want to join the WorkflowHub community?  
See our current activities and upcoming meetings [here](#).

Shortcuts



Discover workflows relating to **SARS-CoV-2 / COVID-19**



**WfCommons**

Looking for WfCommons? [Click here](#)

Current Workflow Types

- Common Workflow Language
- Galaxy
- KNIME
- Nextflow
- Snakemake

...and 24 others

Tags [\[show all\]](#)

Latest additions



Biodiversity Genomics Europe | High Mountain Systems - Arthropod sampling with Malaise traps

SOP - added 6 days ago



AMR-Pathfinder v4.2

Workflow - added 7 days ago



Metaenome-Atlas

Workflow - added 14 days ago



Ocean's variables 2.0

Workflow - added 17 days ago



Evaluation IA-Biodiv workflow

Workflow - added 20 days ago

Find content

Browse Workflows



Browse Collections



Browse Spaces



Browse Teams



WorkflowHub 🔍 Browse ➕ Create  ⚙️ Search 📄 About 📄 Help 📁 My Items 👤 Marie Jossé

Workflow was successfully uploaded and saved.

## Marine Omics identifying biosynthetic gene clusters

Version 1

[📄 Overview](#) [📁 Files](#) [🔗 Related items](#)

**Workflow Type:** Galaxy

Secondary metabolite biosynthetic gene cluster (SMBGC) Annotation using Neural Networks Trained on Interpro Signatures

Click and drag the diagram to pan, double click or use the controls to zoom.



### Version History

**Version 1 (earliest)** Created 21st Aug 2024 at 09:17 by Marie Jossé  
No revision comments

[Open](#) [🔗 master](#) [cd239c1](#)

**Creators and Submitter**

**Creator**  
Marie Jossé

**Submitter**  
 Marie Jossé

---

**Tool**

Prodigal

---

**Citation**

Make your Workflow easily citable by generating a DOI for it.

This Workflow version must be frozen before being eligible for a DOI.

[🔒 Freeze version](#)

# BioDT

Overview **Related items**

## Related items

People (4) **Teams (9)** Organizations (4) SOPs (1) Presentations (1) Workflows (2)

### BioDT additional pipelines



BioDT pipelines not associated with one of the [BioDT use cases](#).

**Space:** BioDT  
**Public web page:** *Not specified*

### BioDT Use Case 4.1.1.1 Biodiversity dynamics



Part of T.4.1.1 *Species response to environmental change* (UFZ, GBIFS, UKCEH, JYU, MLU)

**Space:** BioDT  
**Public web page:** *Not specified*

### BioDT Use Case 4.1.3.1 Invasive species



Part of T.4.1.3 Dynamics and threats from and for species of policy concern (UFZ, GBIFS, Naturalis, UKCI)

**Space:** BioDT  
**Public web page:** *Not specified*

### BioDT Use Case 4.1.1.2 Ecosystem services



## BioDT Guide to using WorkflowHub Version

Overview **Related items**

The [BioDT](#) project is producing and maintaining workflows. We need to register those workflows in WorkflowHub

1. To give visibility to the workflows created by the project and by the different networks and communities
2. To give visibility to the workflows used by project that were created
3. To share workflows across the project, within project networks and externally
4. To credit and cite the people making the workflows and the networks to which they belong
5. To track the versions of workflows as they are produced, and
6. To retain the workflows for post-project use by the networks

This project-specific guide shows how the BioDT will make use of WorkflowHub, such as navigating the BioDT

**SEEK ID:** <https://workflowhub.eu/sops/14?version=1>

**Link:** [https://docs.google.com/document/d/1B-kXTUrmhGy5ZmOfTAxCbN5Tg\\_NIAUYaEdlw8v28hi4/edit](https://docs.google.com/document/d/1B-kXTUrmhGy5ZmOfTAxCbN5Tg_NIAUYaEdlw8v28hi4/edit)

**DOI:** [10.48546/workflowhub.sop.14.1](https://doi.org/10.48546/workflowhub.sop.14.1)

<https://doi.org/10.48546/workflowhub.sop.14.1>  
<https://www.youtube.com/watch?v=380EwriHzHc>

# pDT examples

## Grassland Dynamics - Schrankogel (Austria) - 1994, 2004, 2014

@id	/
name [?]	Grassland Dynamics - Schrankogel (Austria) - 1994, 2004, 2014
@type	Dataset
description [?]	Species cover according to the GLORIA Field Manual: <a href="https://gloria.ac.at/downloads/Manual_5thEd_ENG.pdf">https://gloria.ac.at/downloads/Manual_5thEd_ENG.pdf</a> (p 38); %;1x1m plots arranged in transects at southern slopes of Mt. Schrankogel: <a href="https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.15290">https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.15290</a> ; visual estimate used for the Horizon Europe project "Biodiversity Digital Twin for Advanced Modelling, Simulation and Prediction Capabilities" ( <a href="https://biодt.eu/">https://biодt.eu/</a> ).
author [?]	Christoph Wohner
contentSize [?]	3.9 MB
identifier [?]	<a href="https://doi.org/10.23728/b2share.57315b8581cf45f6a5686b8ec1e0a788">https://doi.org/10.23728/b2share.57315b8581cf45f6a5686b8ec1e0a788</a>
profile	<a href="https://github.com/BioDT/biodt-fair/discussions/8">https://github.com/BioDT/biodt-fair/discussions/8</a>
license [?]	<a href="#">AT_Schrankogel_license.csv</a>
dateCreated [?]	2024-02-20
keywords [?]	<ul style="list-style-type: none"> <li>BioDT</li> <li>Grassland pDT</li> </ul>
url [?]	<a href="https://doi.org/10.23728/b2share.57315b8581cf45f6a5686b8ec1e0a788">https://doi.org/10.23728/b2share.57315b8581cf45f6a5686b8ec1e0a788</a>
variableMeasured [?]	Species cover
measurementTechnique [?]	<a href="#">AT_Schrankogel_license.csv</a>
spatialCoverage [?]	GLORIA Master Site Schrankogel (AT-SCH), Stubaiер Alpen - Austria
temporalCoverage [?]	<ul style="list-style-type: none"> <li>1994</li> <li>2004</li> <li>2014</li> </ul>
version [?]	1
publisher [?]	Long Term Ecological Research Network
hasPart [?]	<ul style="list-style-type: none"> <li><a href="#">AT_Schrankogel_data_cover.csv</a></li> <li><a href="#">AT_Schrankogel_license.csv</a></li> <li><a href="#">AT_Schrankogel_method.csv</a></li> <li><a href="#">AT_Schrankogel_reference.csv</a></li> </ul>

## <https://github.com/BioDT/biodt-fair>

### Cultural Ecosystem Services pDT - Biodiversity Model

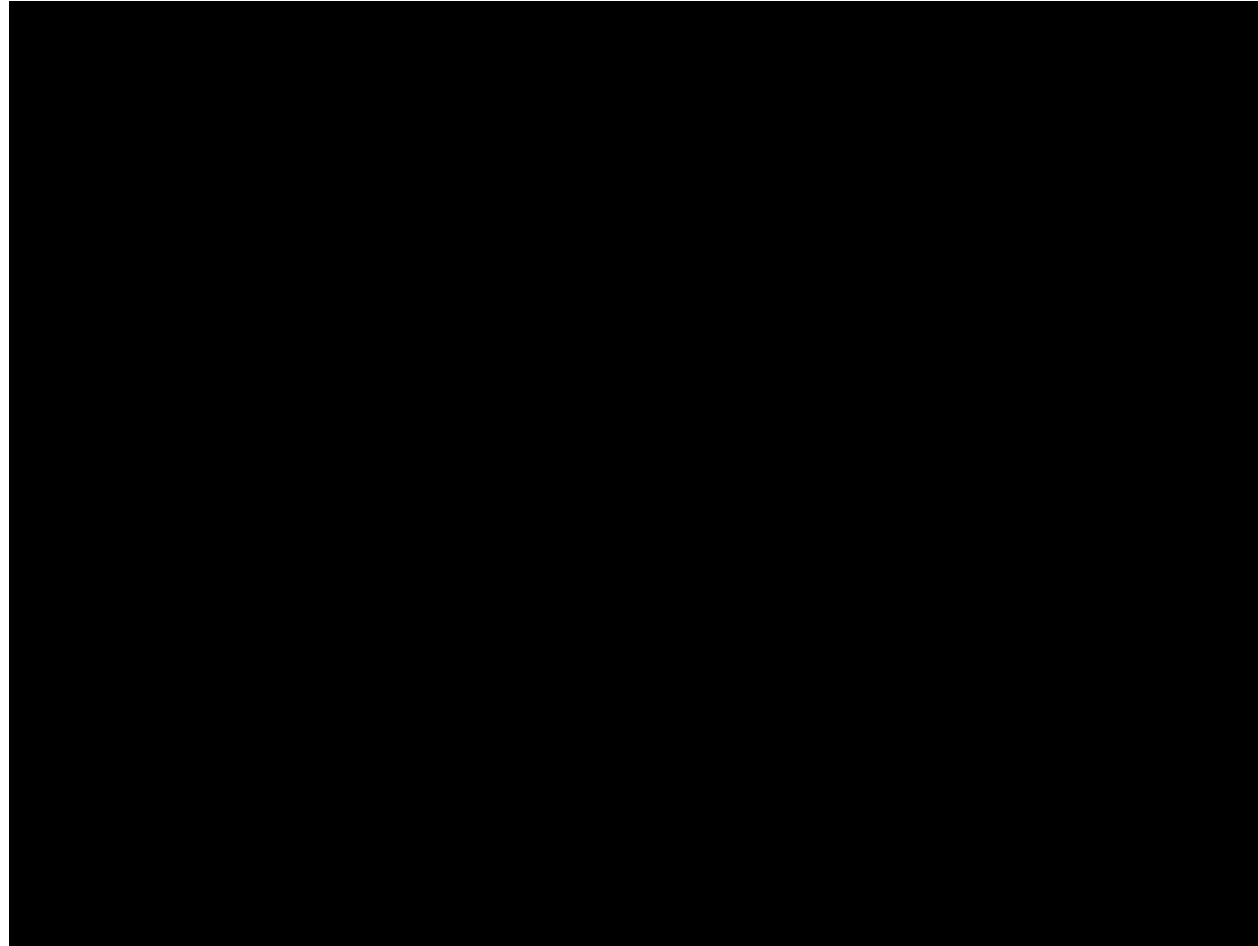

[Download all the metadata for Cultural Ecosystem Services pDT - Biodiversity Model in JSON-LD format](#)  
[Check this crate](#)

### Go to: Species occurrence data from the Cairngorms, Scotland

@id	<a href="https://doi.org/10.15468/dl.b3n3r9">https://doi.org/10.15468/dl.b3n3r9</a>
name	Species occurrence data from the Cairngorms, Scotland
@type	DataFeed
description	Species occurrence data from the Cairngorms, Scotland was obtained by download from GBIF. We filtered environmental variables to only include environment data from within a 5 km buffer of recorded occurrences, and conducted spatial thinning.
encodingFormat	
contentSize	179 KB
variableMeasured	species occurrence
measurementTechnique	
spatialCoverage	<a href="#">Cairngorms National Park</a>
temporalCoverage	
version	
publisher	GBIF
Items that reference this one	
supportingData	<a href="#">Cultural Ecosystem Services pDT - Biodiversity Model</a>

## agroclim workflow demo

agroclim - using crop growth models and DestinE data

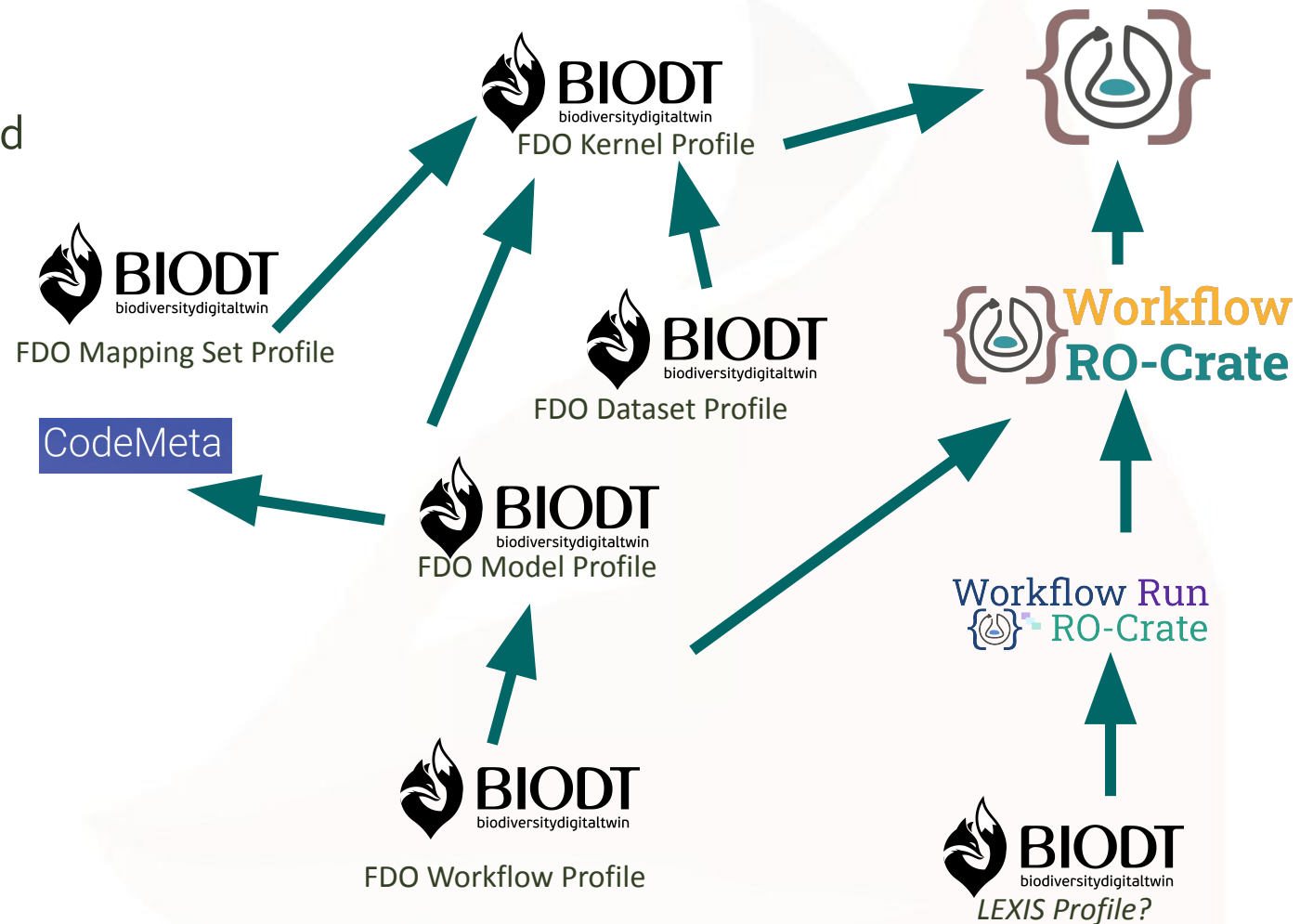


<https://bit.ly/agroclimmov>

# RO-Crate Profiles: Defined Content Checklists

**Profiles** give a set of conventions, types and properties that minimally require and expect to be present in that subset of RO-Crates.

- **Duck typing** for creation, consumption, rendering
- **Classification** for finding and comprehension.
- **Profile Crate** for further defining RO-Crates profile resources
- Key to **extensibility and diversity**



# RO-Crate tutorials & tools

<https://www.researchobject.org/ro-crate/tutorials>

Sandbox (experimental): <https://ro-crate.ldaca.edu.au/>

Crate-O editor: <https://language-research-technology.github.io/crate-o/>

Python: <https://pypi.org/project/rocrate/>

## Creating an RO-Crate

In its simplest form, an RO-Crate is a directory tree with an `ro-crate-metadata.json` file at the top level that contains metadata about the other files and directories, represented by [data entities](#). These metadata consist both of properties of the data entities themselves and of other, non-digital entities called [contextual entities](#) (representing, e.g., a person or an organization).

Suppose Alice and Bob worked on a research task together, which resulted in a manuscript written by both; additionally, Alice prepared a spreadsheet containing the experimental data, which Bob used to generate a diagram. Let's make an RO-Crate to package all this:

```
from rocrate.rocrate import ROcrate

crate = ROcrate()
paper = crate.add_file("exp/paper.pdf", properties={
    "name": "manuscript",
    "encodingFormat": "application/pdf"
})
table = crate.add_file("exp/results.csv", properties={
    "name": "experimental data",
    "encodingFormat": "text/csv"
})
diagram = crate.add_file("exp/diagram.svg", dest_path="images/figure.svg", properties={
    "name": "bar chart",
    "encodingFormat": "image/svg+xml"
})
```

RO-Crate Playground About

Home

Examples

Help

Crate

ro-crate-metadata.json

Other Files

Add Files

Create New

JSON Visual Editor HTML Preview

NEW!! Farms to Freeways Example Dataset  
Current Entity: NEW!! Farms to Freeways Example Dataset

About

Related People, Orgs & Works

Structure

Provenance

Space & Time

Software & Hardware

Others

@id → arcp://name,farms-to-freeways/

Persistent, managed unique ID in URL format (if available), for example a DOI for a collection or an ORCID, personal home page URL or email address for a person

@type → Dataset

RepositoryCollection

+ Select

The type of the entity.

Conforms To → <https://purl.archive.org/language-data-commons/profile#Collection>

Create New Entity

All Entities Unlinked

Enter keyword to filter the val

1 2 3 4 ...

Dataset RepositoryCollection

NEW!! Farms to Freeways Example Dataset

PropertyValue ATAP

Organization

Western Sydney University

ContactPoint

Contact Katrina Trewin

OrganizationReuseLicense

Attribution 3.0 Australia (CC BY 3.0 AU)

Validate

SUCCESS: A JSON-LD document that describes the RO-Crate with structured data in the form of RO-Crate JSON-LD.

SUCCESS: Has a context <https://w3id.org/ro-crate/1.1/context> named "RO-Crate JSON-LD Context", version 1.1.3

WARNING: Root Data Entity has appropriate @id. Is: <arcp://name,farms-to-freeways/corpus>

SUCCESS: Found required property: @type

SUCCESS: Found required property: @id

SUCCESS: Found required property: name

SUCCESS: Found required property: description

SUCCESS: Found required property: datePublished

## Schedule

	Setup	Download files required for the lesson
00:00	1. Introduction	How do I package data in a FAIR way? How can I list the authors of individual files? Can I use multiple licenses in the same data package? How can I visualize JSON-LD metadata?
00:17	2. Turning a folder into an RO-Crate	How can I start a new RO-Crate?
00:24	3. Making a metadata descriptor	Which RO-Crate version is used? How can the crate self-identify as an RO-Crate?
00:28	4. Declaring the root folder	What is the root folder?
00:31	5. Describing the root entity	How can I describe the crate? How do I specify the license of the RO-Crate?
00:39	6. Adding cross-references	How can I describe an entity further? How can I cross-reference different entities?
00:46	7. Data entities	How do I describe the files in my RO-Crate?
00:50	8. Contextual entities	How can I describe things in the world? How can I give details about licenses?
00:54	9. Authorship in crates	How can I list who made the content of the crate? How do I affiliate a person with their place of work?
01:00	10. Validating JSON-LD	How can I validate the JSON-LD?



## Current RO-Crate trends in biodiversity & earth sciences

RO-Crate and Signposting “[Webby FDOs](#)” is recognized as a viable **FAIR Digital Objects** implementation (FDO2024)

**Biodiversity Genomics Europe** using RO-Crate for ENA submissions

**Dataverse** has added RO-Crate support

RO-Crate in **climate science**: FAIR-EASE, CLIMATE-ADAPT4EOSC, FAIR2ADAPT

Digital Twin for GEOphysical extremes (**DT-GEO**): RO-Crate both for HPC workflows (COMPSs) and in EPOS Metadata schema

## Next steps (a taster!)

**EuroScienceGateway:** Workflows as FAIR Digital Objects in EOSC

**EOSC-ENTRUST, EVERSE and HDR UK** using RO-Crate for federated analytics  
“Five Safes RO-Crate”

**BioFAIR UK**: BioCommons infrastructure for UK life science researchers

**NFDI4DS:** Hackathon 2024: [Webby FDOs with RO-Crates and FAIR](#)

[Signposting](#)

**RO-Crate governance:** Community driven (2018→)

→ Steering committee (2025)

### Minutes/agenda

[2024-10-24 Research Object Crate #98](#)

[2024-10-10 Research Object Crate #97 \(EOSC\)](#)

[2024-09-26 Research Object Crate #96](#)

[2024-09-12 Research Object Crate #95 \(EOSC/governance\)](#)

# Recent publications

Otso Ovaskainen, et al. (2024): **Prototype Biodiversity Digital Twin: Real-time bird monitoring with citizen-science data.** *Research Ideas and Outcomes* 10:e125523

<https://doi.org/10.3897/rio.10.e125523>

Simon Rolph, et al. (2024): **Prototype Digital Twin: Recreation and biodiversity cultural ecosystem services.** *Research Ideas and Outcomes* 10:e125450 <https://doi.org/10.3897/rio.10.e125450>

Desalegn Chala, et al. (2024): **Prototype biodiversity digital twin: crop wild relatives genetic resources for food security.** *Research Ideas and Outcomes* 10:e125192

<https://doi.org/10.3897/rio.10.e125192>

Jonas Grieb, et al. (2024): **Using Webby FDOs to Integrate AI Taxon Identification and Citizen Science.** *Biodiversity Information Science and Standards* 8:e134757 <https://doi.org/10.3897/biss.8.134757>

Claus Weiland, et al. (2024): **Dataspace Integration for Agrobiodiversity Digital Twins with RO-Crate.** *Biodiversity Information Science and Standards* 8:e134479 <https://doi.org/10.3897/biss.8.134479>

Tomáš Martinovič, et al. (2024): **Graphical User Interface for Biodiversity Digital Twins: Data Challenges.** *Biodiversity Information Science and Standards* 8:e135092

<https://doi.org/10.3897/biss.8.135092>

Sharif Islam, et al. (2024): **Bridging Data Standards and FAIR Principles in Biodiversity Digital Twinning: Prototyping, Challenges, Lessons Learned, and Future Plans.**

*Biodiversity Information Science and Standards* 8:133089 <https://doi.org/10.3897/biss.8.133089>

Rafael Ferreira da Silva, et al. (2024):

**Workflows Community Summit 2024: Future Trends and Challenges in Scientific Workflows.** <https://doi.org/10.5281/zenodo.13844759>

Ove Johan Ragnar Gustafsson, et al. (2024): **WorkflowHub: a registry for computational workflows.** *arXiv:2410.06941* [cs.DL]

<https://doi.org/10.48550/arXiv.2410.06941>

Sean Wilkinson, et al. (2024): **Applying the FAIR Principles to Computational Workflows.** *arXiv* 2410.03490 [cs.DL]

<https://doi.org/10.48550/arXiv.2410.03490>

Simone Leo, et al. (2024): **Recording provenance of workflow runs with RO-Crate.** *PLOS One* 19(9):e0309210

<https://doi.org/10.1371/journal.pone.0309210>

Stian Soiland-Reyes, et al. (2024): **Practical webby FDOs with RO-Crate and FAIR Signposting: Experiences and lessons learned.** International FAIR Digital Objects Implementation Summit 2024 (FDOF 2024). <https://s11.no/2024/webby-fdos/>



**BIODT**  
biodiversitydigitaltwin

 @BiodiversityDT

 BioDT



Funded by  
the European Union