



**From publication to reuse:
Making data in Dryad
data science friendly**



An **open data publishing platform**
& **community** committed to the
open availability and routine
re-use of all research data

- **Serving all research domains**
- **Leader in research data**
- **Open source and interconnected**
- **Fully curated**



DRYAD

Not-for-profit and online since
2008. Learn more: datadryad.org

“Data should be of sufficient quality to validate and replicate research findings”

*

https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html#_ftn8



DRYAD

Not-for-profit and online since
2008. Learn more: datadryad.org

A close-up photograph of a hand holding a yellow pencil vertically next to a cluster of small, vibrant purple flowers with white centers. The flowers are growing on a dark, rocky surface. The hand is positioned in the lower-left corner, and the pencil is held between the thumb and index finger. The background shows more of the same flowers and some green leaves with serrated edges. The overall scene is brightly lit, highlighting the colors of the flowers and the texture of the rock.

Enabling data reuse— a holistic approach

Improve dataset descriptions

```
# Title of Dataset:
```

```
---
```

```
Brief summary of dataset contents, contextualized in experimental procedures  
and results.
```

```
## Description of the Data and file structure
```

```
This is a freeform section for you to describe how the data are structured  
and how a potential consumer might use them. Be as descriptive as necessary.  
Keep in mind that users of your data might be new to the field and unfamiliar  
with common terminology, metrics, etc.
```

```
Describe relationship between data files, missing data codes, other  
abbreviations used. Be as descriptive as possible.
```

```
## Sharing/access Information
```

```
Links to other publicly accessible locations of the data:
```

```
Was data derived from another source?
```

```
If yes, list source(s):
```



DRYAD

Not-for-profit and online since
2008. Learn more: datadryad.org

Employ standards and persistent identifiers



Character set and phylogenetic analyses of the living and fossil egeriine scincids of Australia

Thorn, Kailah, University of Western Australia, <https://orcid.org/0000-0002-0645-835X>

Hutchinson, Mark, South Australian Museum

Lee, Michael, Flinders University

Camens, Aaron, Flinders University

Worthy, Trevor, Flinders University, <https://orcid.org/0000-0001-7047-4680>

kailah.thorn@uwa.edu.au, mark.hutchinson@samuseum.sa.gov.au, mike.lee@flinders.edu.au, aaron.camens@flinders.edu.au, trevor.worthy@flinders.edu.au

Publication date: February 10, 2021

Publisher: Dryad

<https://doi.org/10.5061/dryad.3n5tb2rg7>

Citation

Thorn, Kailah et al. (2021), Character set and phylogenetic analyses of the living and fossil egeriine scincids of Australia, Dryad, Dataset, <https://doi.org/10.5061/dryad.3n5tb2rg7>

Abstract

Data files

[Download dataset](#)

> February 4, 2021
> February 10, 2021

Metrics

98 views

243 downloads

0 citations

Keywords

Biological sciences

Persistent identifiers “play a role in the reusability of data by enabling rich metadata and provenance to be associated with a digital object.”

Cousijn, et al “[Connected Research: The Potential of the PID Graph](https://doi.org/10.1016/j.patter.2020.100180)”, <https://doi.org/10.1016/j.patter.2020.100180>



DRYAD

Not-for-profit and online since 2008. Learn more: datadryad.org

Link to related artifacts

Related Works

Are there any preprints, articles, datasets, software packages, or supplemental information that have resulted from or are related to this Data Publication?

Article	https://doi.org/10.1371/journal.pone.0238882	remove
Software	10.5281/zenodo.2583172	remove
Dataset	http://purl.org/phylo/treebase/phyloWS/study/TB2:S1929	remove

[+ add another related work](#)



Characterizing and through microRNA

Nanayakkara, Jina, Queen's Unive
Yang, Xiaojing, Queen's University
Tyryshkin, Kathrin, Queen's Unive
Wong, Justin J.M., Queen's Univer
Vanderbeck, Kaitlin, Queen's Univ
Ginter, Paula S., Weill Cornell Med
Scognamiglio, Theresa, Weill Cor

Works Referencing This Dataset

Panarelli, Nicole et al. (2019), Evaluating gastroenteropancreatic neuroendocrine tumors through microRNA sequencing, Endocrine-Related Cancer, Journal-article, <https://doi.org/10.1530/erc-18-0244>

Renwick, Neil et al. (2013), Multicolor microRNA FISH effectively differentiates tumor types, Journal of Clinical Investigation, Journal-article, <https://doi.org/10.1172/jci68760>

Cheung, Irene Y. et al. (2014), Deep MicroRNA sequencing reveals downregulation of miR-29a in neuroblastoma central nervous system metastasis, Genes, Chromosomes and Cancer, Journal-article, <https://doi.org/10.1002/gcc.22189>



DRYAD

Not-for-profit and online since
2008. Learn more: datadryad.org

Improve data quality

Upload Your Files

You may upload data via two mechanisms: directly from your computer, or from a URL on an external server (e.g., Box, Dropbox, AWS, lab server). We do not recommend using Google Drive.

We require that you include a README file to provide key information for understanding and using your data.

Software and Supplemental Information can be uploaded for publication at [Zenodo](#). You will have the opportunity to choose a separate license for your software on the review page.

 Data e.g., csv, fasta Choose Files Enter URLs	 Software e.g., code packages, scripts Choose Files Enter URLs	 Supplemental Information e.g., figures, supporting tables Choose Files Enter URLs
--	--	--

Files

Filename	Status	Tabular Data Check	URL	Type	Size	Actions
file_example_XLS_100.xls	Uploaded	Passed		Data	20.48 KB	Remove
dataset-analysis-output_reports.Rmd	Uploaded			Software	4.35 KB	Remove
file_example_XLS_5000.xls	<input checked="" type="checkbox"/>			Data	672.26 KB	Remove

I confirm that no Personal Health Information or Sensitive Data are being uploaded with this submission.

[Upload pending files](#)



DRYAD

Not-for-profit and online since
2008. Learn more: [datadryad.org](#)



**FRICTIONLESS
DATA**

Provide data previews

The screenshot shows the DRYAD website interface. At the top left is the DRYAD logo. To the right is a search bar and navigation links: "Explore data", "About", "Help", and "Login". The main title of the dataset is "Data from: The SCOC database – a large, open and global database with sediment community oxygen consumption rates". Below the title, the author information is listed: "Stratmann, Tanja, Royal Nethelands Institute for Research and Innovation Soetaert, Karline Wei, Chih-Lin Lin, Yu-Shih van Oevelen, Dick tanja.Stratmann@nioz.nl". The publication date is "October 4, 2022" and the publisher is "Dryad". A DOI link is provided: "https://doi.org/10.5061/dryad.25nd083".

A "Data files" section on the right lists two files: "List_of_studies_version_1.csv" (69.94 kB) and "SCOC_database_version_1.csv" (1.23 MB). A "Download dataset" button is visible above the list.

An overlay window titled "CSV preview of List_of_studies_version_1.csv" is centered on the page. It contains a table with columns: "Study reference", "Screening phase", "Eligibility phase", and "SCOC database". The first row of data is: "Abhilash, K. R., Raveendran, T. V., Limma Mol, V. P. & Deepak, M. P. 2015, 2019". The second row is: "Accornero, A. et al. Organic carbon budget at the sediment-water interface in the Amazon estuary, Brazil". The third row is: "Ahmerkamp, S. et al. Regulation of benthic oxygen fluxes in permeable sediments under varying oxygen concentrations". The fourth row is: "Algeo, T. J. & Heckel, P. H. The late Pennsylvanian midcontinent sea level rise: Evidence from the Antrim Shale, Michigan". The fifth row is: "Alkhatib, M., del Giorgio, P. A., Gelin, Y. & Lehmann, M. F. Benthic oxygen consumption in a temperate estuary". The preview is truncated with "..." at the bottom.

Below the preview is a button labeled "Copy preview text to clipboard".

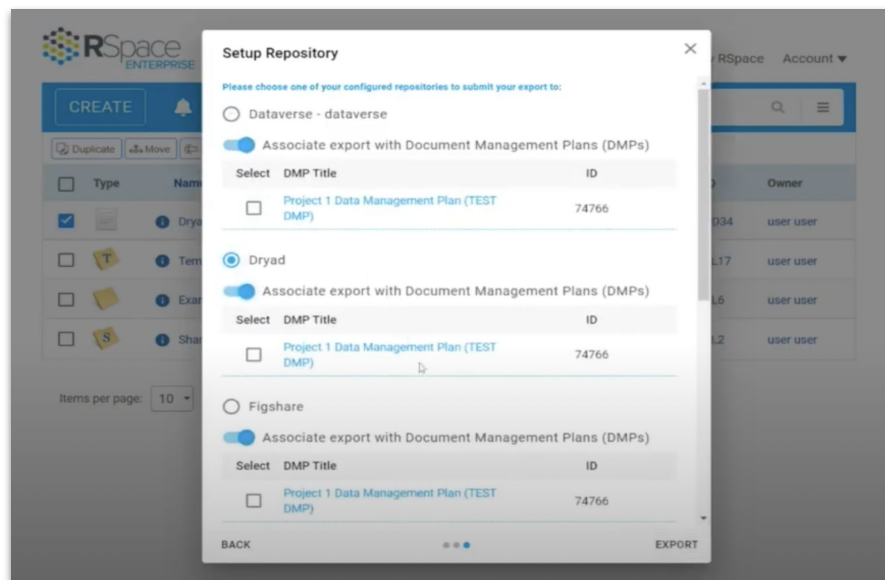
At the bottom of the page, there are statistics: "285 views", "132 downloads", and "0 citations".



DRYAD

Not-for-profit and online since 2008. Learn more: datadryad.org

Connect with open science tools



<https://www.youtube.com/watch?v=T4TjcjYv1ic>



DRYAD

Not-for-profit and online since 2008. Learn more: datadryad.org



Enable machine access

dryad-api **2.1.0** **OAS3**

[/openapi.yml](#)

The API for the [Dryad](#) data publication and preservation platform.

In addition to this documentation, please see the [descriptive API documentation](#), particularly the [API Submission Examples](#) which give concrete examples of submission through the Dryad API.

Our [Github documentation directory](#) will also be expanded further in the future as more documentation is created (though not all of it may be relevant to API users).

Rate limits:

- Anonymous users of the API are limited to 30 requests per minute, with a lower limit for downloads of data files
- Authenticated users are limited to 100 requests per minute

rdryad

repo status **Active** R-check **failing** codecov **67%** downloads **171/month** CRAN **1.0.0**

rdryad is a package to interface with the Dryad data repository.

General Dryad API documentation: <https://datadryad.org/api/v2/docs/>

rdryad docs: <https://docs.ropensci.org/rdryad/>

Installation

Install Dryad from CRAN



DRYAD

Not-for-profit and online since
2008. Learn more: datadryad.org

License for reuse

License

This work is licensed under a [CC0 1.0 Universal \(CC0 1.0\) Public Domain Dedication](#) license.



DRYAD

Not-for-profit and online since 2008. Learn more: datadryad.org

Coming soon



Use open formats

Obstacle

Dryad authors often use Excel sheets to combine datasets. This is challenging because consumers cannot easily understand what datasets are included in a submission.

Solution

Export Excel sheets to individual CSV files upon submission before running validation. Use the sheet names as the names of the CSV files.

Benefit

A Dryad consumer can easily see a list of all datasets via the API on the web and also programmatically access individual files within and across submissions.



DRYAD

Not-for-profit and online since
2008. Learn more: datadryad.org

Increase visibility into data packages

Obstacle

When viewing a submission, a reader cannot see how many files it contains, especially if it is an opaque zip, tar, or gz.

Solution

Provide this information to the user via the interface and the API.

Benefit

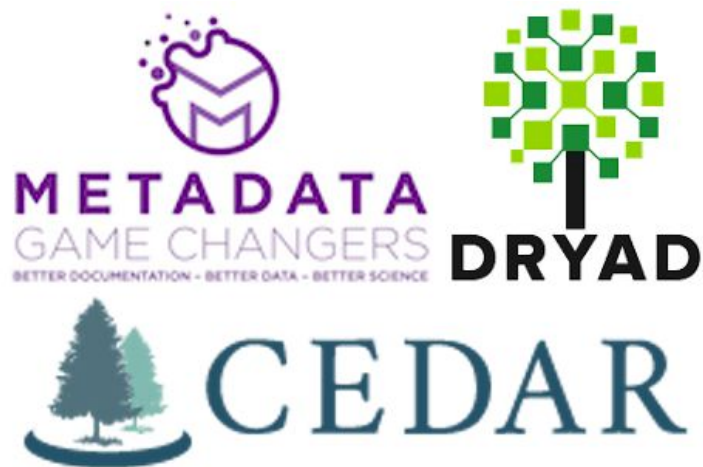
Users can quickly see how many and what files they are dealing with before making a download request. Users who query the API directly or via high-level packages (written in R or Python) can have reproducible workflows.



DRYAD

Not-for-profit and online since
2008. Learn more: datadryad.org

Improve
dataset
quality



DRYAD

Not-for-profit and online since
2008. Learn more: datadryad.org

Stay up to date

→ github.com/CDL-Dryad/dryad-product-roadmap

→ blog.datadryad.org/



DRYAD

Not-for-profit and online since
2008. Learn more: datadryad.org

Contact us

Help & support help@datadryad.org

General inquiries hello@datadryad.org

Community sarah@datadryad.org
