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

Sarah Lippincott, Head of Community Engagement and Maria Praetzellis, Interim Product Manager
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
“Data should be of sufficient quality to validate and replicate research findings”

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

Related Works

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

<table>
<thead>
<tr>
<th>Type</th>
<th>URI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article</td>
<td><a href="https://doi.org/10.1371/journal.pone.0238882">https://doi.org/10.1371/journal.pone.0238882</a></td>
</tr>
<tr>
<td>Software</td>
<td>10.5281/zenodo.2583172</td>
</tr>
<tr>
<td>Dataset</td>
<td><a href="http://purl.org/phylo/treebase/phyllows/study/TB2:S1929">http://purl.org/phylo/treebase/phyllows/study/TB2:S1929</a></td>
</tr>
</tbody>
</table>

+ add another related work

Works Referencing This Dataset


Link to related artifacts
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.

Choose Files

Files

<table>
<thead>
<tr>
<th>Filename</th>
<th>Status</th>
<th>Tabular Data Check</th>
<th>URL</th>
<th>Type</th>
<th>Size</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>file_example_XLS_100.xls</td>
<td>Uploaded</td>
<td>Passed</td>
<td></td>
<td>Data</td>
<td>20.48 KB</td>
<td>Remove</td>
</tr>
<tr>
<td>dataset-analysis-output_reports.Rmd</td>
<td>Uploaded</td>
<td></td>
<td></td>
<td>Software</td>
<td>4.35 KB</td>
<td>Remove</td>
</tr>
<tr>
<td>file_example_XLS_5000.xls</td>
<td></td>
<td></td>
<td></td>
<td>Data</td>
<td>672.26 KB</td>
<td>Remove</td>
</tr>
</tbody>
</table>

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

Upload pending files
Provide data previews
Connect with open science tools

https://www.youtube.com/watch?v=T4TjcjYv1ic
Enable machine access

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 file
- Authenticated users may attempt to download a single file at a time

rdryad

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
License for reuse

License

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

GitHub is Sued, and We May Learn Something About Creative Commons Licensing
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.
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.
Improve dataset quality
Stay up to date

➔ github.com/CDL-Dryad/dryad-product-roadmap
➔ blog.datadryad.org/
Contact us

Help & support  help@datadryad.org
General inquiries  hello@datadryad.org
Community  sarah@datadryad.org