



CONNECTING WITH RESEARCHERS VIA LOCAL DATA CURATION: THE UCSB EXPERIENCE WITH DRYAD DEPOSITS

Greg Janee & Renata Curty

Research Data Services

UC SANTA BARBARA

Library





Self-archiving

Disciplinary Agnostic

Integration with various academic journals

Integration with Zenodo

DOIs

CC0 License...

Curation of Deposits (FAIR principles)

How did it begin?



JUNE, 2020

We became part of the Data Curation Network (DCN)

Check files
Understand or try to
Request missing information
Augment the submission
Transform the format
Evaluate for FAIRness
Document throughout

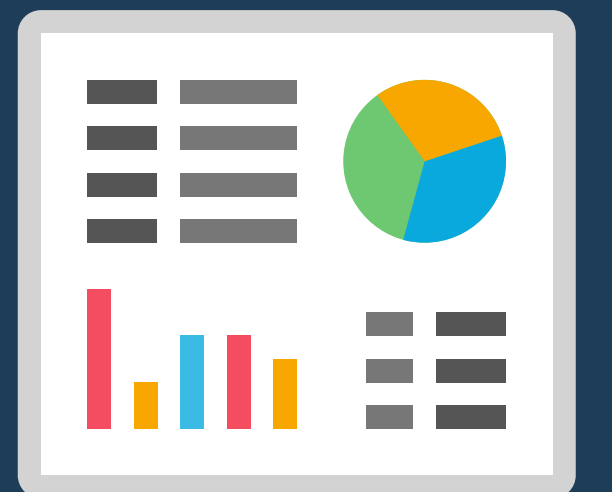
MARCH, 2021

Dryad Curation Onboarding



APRIL, 2021

Internal workflow and ticketing system (Jira)



Admin Dashboard

At a glance 76,605 Users 63,712 Datasets	Activity in the last 7 days 0 users added 1 datasets started 2 datasets submitted
---	---

Datasets

Search Terms: All terms [Clear Search](#)

Filter by: Status:

Journal Name:

[Reset all filters](#)

Title	Status	Author	DOI	Last Modified	Curator	Size	Publication Date
The CALFISH database: open-sourcing a century of California's non-confidential fisheries landings and participation data	Published	Bellquist; Free; Vargas Poulsen	10.25349/D9M907	04/16/2021 19:29:48 UTC	Renata Gonçalves Curty	7.35 MB	April 16, 2021
Data from: Characterizing driver-response relationships in marine pelagic ecosystems for improved ocean management	Published	Amrhein; Halpern; Hunsicker; Kappel; Mease; Scarbo ...	10.5061/dryad.k5n98	06/24/2020 19:45:57 UTC		157.20 kB	October 15, 2015
Data from: Regulation of reproductive processes with Dynamic Energy Budgets	Published	Casas; Gergs; Lika; Muller; Murphy; Nacci; Nisbet; ...	10.5061/dryad.58j9r88	06/24/2020 18:38:57 UTC		31.38 kB	February 2, 2019
Data from: Evolutionary relatedness does not predict competition and co-occurrence in natural or experimental communities of green algae	Published	Alexandrou; Bentlage; Cardinale; Delwiche; Fritsch ...	10.5061/dryad.c574h	06/30/2020 17:20:40 UTC		8.27 MB	November 14, 2014
Data from: A mid-Oligocene (Whitneyan) rhinocerotid from northeastern California	Published	Bright; Tiffney; Wyss	10.5061/dryad.dt682	06/24/2020 19:26:18 UTC		23.14 MB	February 16, 2015
Data from: Phylogenetic	Published	Allen; Ault; Betancourt;	10.5061/dryad.td03n886	06/24/2020		1.15	April 2, 2021

- Dryad curation doi:10.253...
- RDS-192 Dryad curation doi: 10.253...
- RDS-214 Dryad curation doi:10.253...
- RDS-170 Dryad curation doi:10.253...
- RDS-212 Dryad curation doi:10.253...

Research Data Services / RDS-170 13 of 54

Dryad curation doi:10.25349/D9TW33

Details

Type: Curation Status: RESOLVED

Priority: Normal Resolution: Done

Labels: None Security Level: Project Users (All project Users can view the issue)

Dataset Name: Surgeonfish foraging data at Palmyra Atoll

Dataset DOI: 10.25349/D9TW33

Depositor Name: Jacob G. Eurich

Position: Researcher

Affiliation: Marine Science

Data Types: Excel/Tabular

Curation Status: Published

Actions Requested from Depositor: Add a README

Curation Actions Performed: Add missing documentation, Edit data for accuracy, Enhance record metadata, ... (3)

People

Assignee: Renata Curty

Reporter: Renata Curty

Watchers: Stop watching this issue

Dates

Created: 04/15/2021 08:57 AM

Updated: 04/29/2021 11:21 AM

Resolved: 04/29/2021 11:14 AM

Development

[Create branch](#)

Description

P.S. Laura is the first author, but not the depositor. Jacob (affiliated requested to be listed as the first author.

Dear Laura E. Dee,

Thank you for submitting your dataset "Surgeonfish foraging data at Palmyra Atoll" (doi:10.25349/D9TW33) to Dryad.

Your dataset will be curated shortly, and the below are the next steps:

We encourage you to check that your files, metadata, contributor information and relevant documentation adhere to our submission guidelines. [Curation checklist]. Assuring that your dataset meets all of our requirements will ensure that the curation process is as efficient and timely as possible.

Your submission can be viewed and revised throughout the process through 'My Datasets' under 'Submitted'.

You will receive an email if the curators have questions about your dataset or upon publication if there are no issues.

Your dataset has been assigned a unique identifier, called a DOI (doi:10.25349/D9TW33). If your dataset is associated with a manuscript submission, you may provide this DOI to the journal, although it will not be live until the dataset is published.

How is it going?

Curation Activity

- 26 curated deposits
- 7 in progress
- 21 awaiting associated manuscript peer-review
- Predominantly from Marine Sciences, Environmental Sciences, and Biology
- Tabular data is the most common format



Most frequent curation actions requested or performed

- Add or Edit existing README file
- Add or Edit supporting documentation
- Enhance metadata



"Thank you for all of the support!"

Postdoctoral Researcher

"I really appreciate the work that you're doing to help faculty and students archive our data!"

Faculty

"Thanks so much for the thorough review of the datafiles".

PhD Candidate

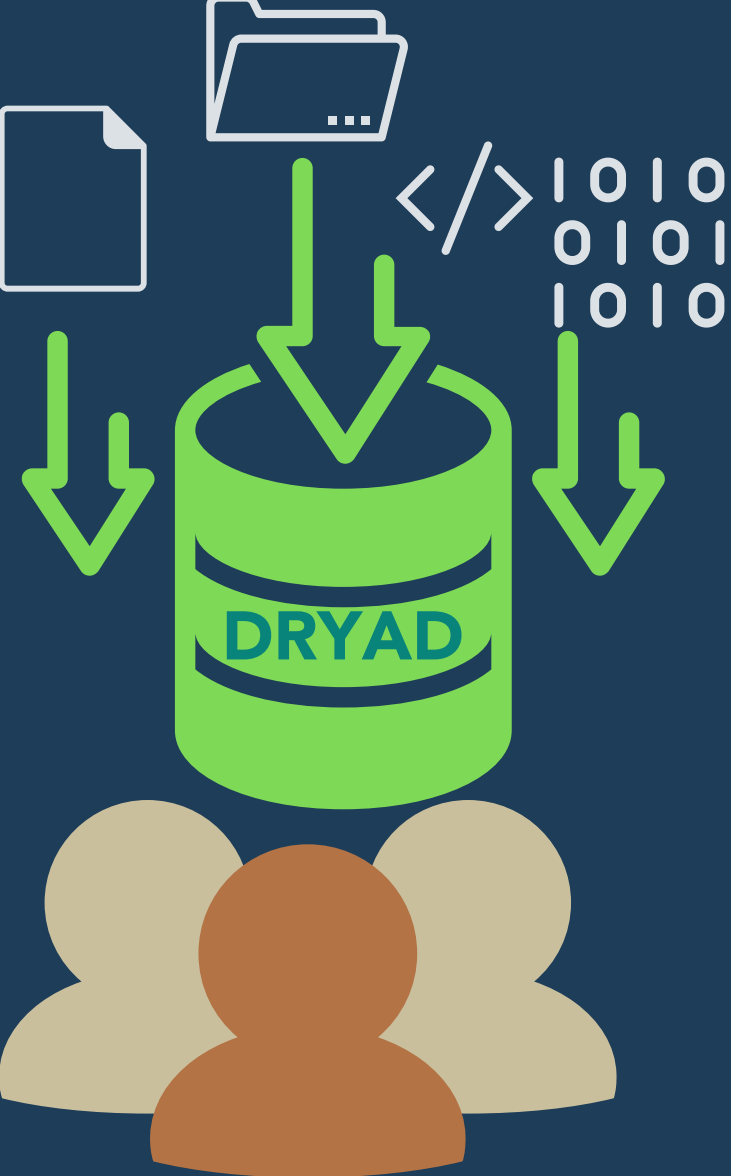
Beyond Curation...



UCSB
Researchers



Research
Workflow



RDS
Team

Establishing Connections | Data Literacy & Education | Instructional Materials

Example

Research Data Services Data Literacy Series, n.3/2021

README, please!

A README is a simple text master file that should include information and metadata about other files in a directory. It should describe the underlying data of your project, and how files are logically associated with one another, in order to make them interpretable and reusable.

How to organize your README

- 1. General information**
Include information about the provenance of the dataset
- 2. Data and file overview**
Describe all files and how they relate to each other
- 3. Sharing and accessing information**
Add licenses or restrictions, links to related-publications and other publicly accessible locations, and the recommended citation for the data
- 4. Methodological information**
Specify the methods adopted to collect or generate and processing the data
- 5. Data-specific information**
List all variables, units of measurement, codes, and symbols used to record data and missing data

TOP TIPS!

- Write your README as a plain text file
- Use standardized formats
- Follow existing scientific conventions for your discipline
- Place the README in the same directory as the data files

Have a question or want some feedback on your README?
rds@library.ucsb.edu

Scan the QR code for a customizable template!
<http://tiny.cc/README>

UC SANTA BARBARA Library
www.library.ucsb.edu

Research Data Services Data Literacy Series, n.08/2021

What should you keep in your project folder?

Over the course of your project, you might produce multiple files. To ease data management and make your research more shareable and reproducible, we suggest you organize your project folder following this basic structure:

Your-Project

/Data

- ./InputData
Often not easy to re-obtain. Supports inspection/backtracking.
- ./IntermediateData
Stores all processes (manipulations and transformations) performed with the input/raw data.
- ./AnalysisData
The "analysis-ready" version of the datasets.

/Scripts

- ./Master-Scripts
If you have multiple scripts, provide the master scripts that call all other scripts used for processing, analyzing the data, and producing visualizations in sequence.
- ./Processing-Scripts
- ./Analysis-Scripts
- ./Visualization-Scripts
Make sure your script/code is:
1. easy to understand,
2. efficient when run, and
3. well-documented.

/Results

- ./Visualization-Files
Should include all output files that capture results generated by the analysis scripts.
- ./Tables
- ./Graphs
- ./Log-Files
- ./Text-Files
Highlight in the output file(s) the numbers, figures, or charts used in the associated reports/manuscripts.

Report.pdf

README.txt
Click here for more info & template

Depending on the nature of your project, you may also create additional subfolders to store supporting documentation such as data collection instruments, IRB protocols, etc.

QUICK TIPS

- Follow best practices for file naming (see tips here).
- Use a version control system such as Git to manage changes in files over time.
- Choose open and preservation-friendly file formats (see tips here).

Have a question or want to schedule a consultation with us? rds@library.ucsb.edu

UC SANTA BARBARA Library
www.library.ucsb.edu

Some of the most frequent opportunities for improvement in data deposits have been covered in our Data Literacy Series

LOOKING AHEAD



1

Promote best practices for code and software sharing

2

End-user Assessment

3

Host a local Dataverse

4

Increase deposits from the Humanities and Social Sciences