

## Hi, I'm Sarah!

Head of community engagement at Dryad

- Librarian
- Open data enthusiast
- Here to help!

sarah@datadryad.org help@datadryad.org



# What you'll learn today

- When (and why) Dryad a good choice for your data
- What to do to prepare your data for sharing
- How to submit to Dryad and what happens along the way
- How Dryad helps extend the reach of your data and how you can keep track of engagement



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
- Interconnected

49,600+ data publications 193,300+ researchers 69,100+ international institutions 1,270+ academic journals



## **Basics**

### Data and metadata are:

Fully curated

Permanently stored in our CoreTrustSeal repository

Published under a Creative Commons Public Domain (CCo) license

Accessible via our open API





### What data is a best fit for Dryad?

### **Great fit**

- Data from any discipline (and interdisciplinary data)
- Data ready and intended for broad sharing and reuse

### Has a better home elsewhere

- Data that has an appropriate home in a disciplinary or specialist repository
- Data that requires a long-term embargo or managed access
- Data containing personally identifiable information (PII) or other sensitive content



## Any field. Any format.

Submit data in any file format from any field of research. Share all of the data from a project in one place.



## Quality control and assistance.

Our curators will check your files before they are released, and help you follow best practices.



## **Engagement and impact.**

See the impact of your data through metrics and citations. We assign your dataset a DOI and make it discoverable online.



## Straightforward compliance.

Submit your data to satisfy publisher and funder requirements for preservation and availability with a minimum of effort.



## Community-led.

Dryad is a nonprofit membership organization that is committed to making data available for research and educational reuse now and into the future.



## Sponsored.

Through your institution's Dryad membership, data publication charges are 100% covered.



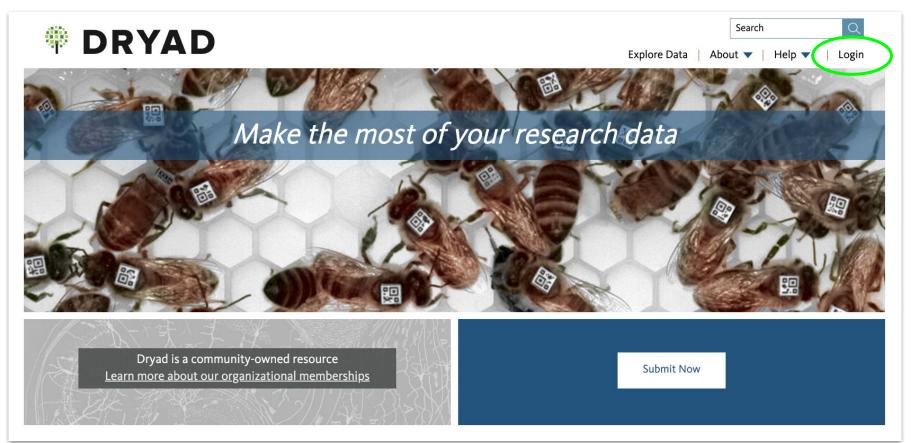
## Access

Authentication via ORCID

Affiliation verification through SSO







### datadryad.org



### **ORCID Account**

Dryad uses ORCID as its primary login method. Please use, or create, your ORCID login credentials to login to Dryad.



### Dryad's Commitment to You

### Curation

All datasets are curated to ensure they are Findable, Accessible, Interoperable, and Reusable

### Compliance

Enabling adherence to funder and publisher open data mandates

### **Community Supported**

Dryad is researcher-led and supported by our institutional and publisher members





### Login

ORCID login successful

Your institution may be a member\* of Dryad.

### verify Cal State East Bay Claremont College Services (TCCS) Claremont Graduate University Claremont McKenna College Columbia University Desert Research Institute Harvey Mudd College Indian Institute of Technology Dhanbad Keck Graduate Institute se Dryad. King Abdullah University of Science and Technology **Kyoto University** Lawrence Berkeley Lab Priva Macquarie University Montana State University Copy Nederlands Instituut voor Ecologie North Carolina State University

## Login Username: Password: Log in For security reasons, please log out and exit your web browser when you have finished accessing services that require authentication.



## Preparation

Gather all data needed for reanalysis

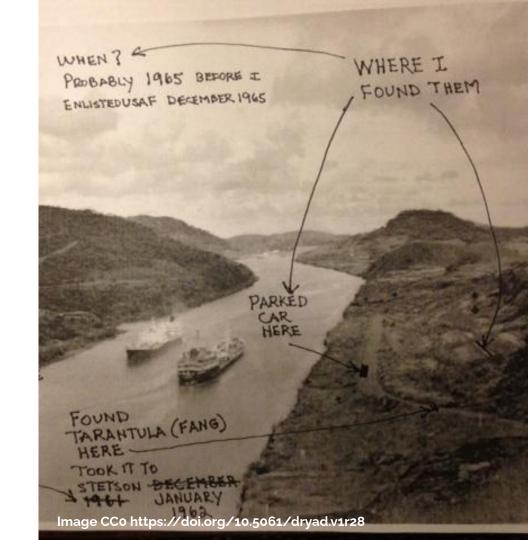
Verify files can be shared publicly

Choose open file formats

Organize files logically

Describe your dataset in a detailed README file





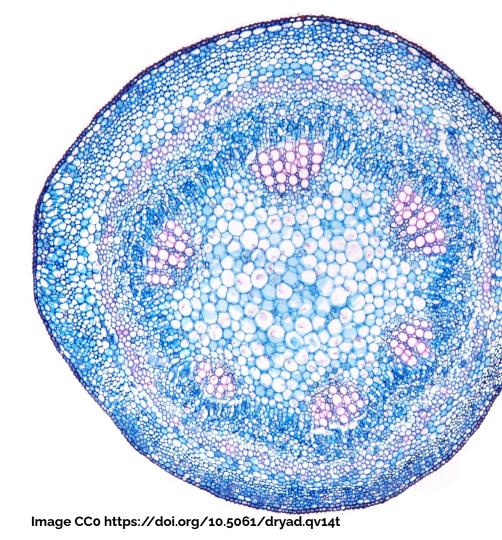
## Submission

Collects necessary metadata

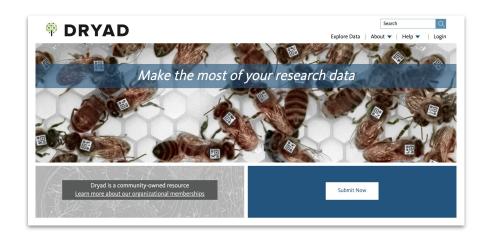
Links dataset to other outputs

Ingests and validates files





### **Pathways**







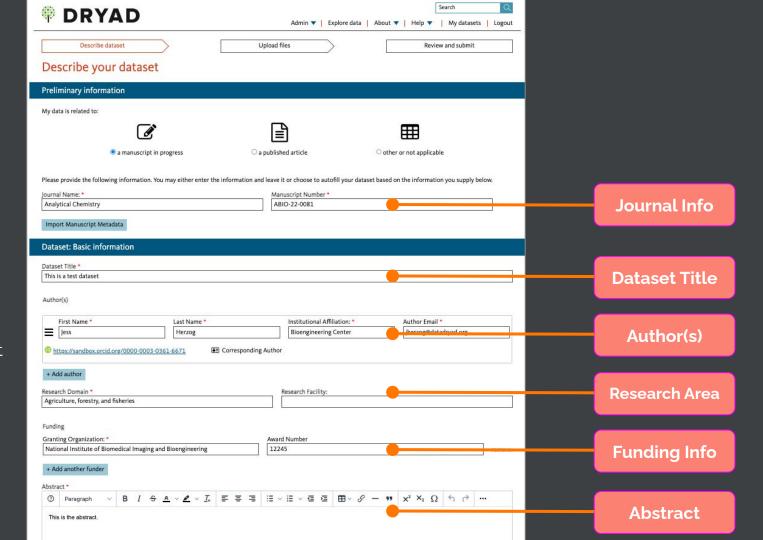


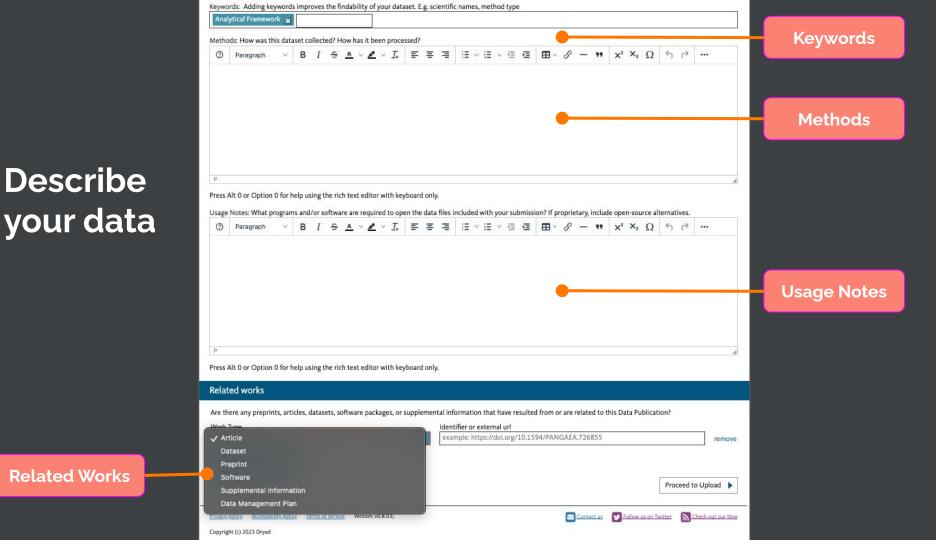


https://blog.datadryad.org/2023/01/13/for-journals-integrate-with-dryad-now-with-ejournalpress/https://blog.datadryad.org/2021/10/21/product-update-editorial-manager-integrates-with-dryad/https://www.youtube.com/watch?v=T4TjcjYv1ic

## Describe your data

**DOI assigned** upon submission & does not change





### Uploac your da

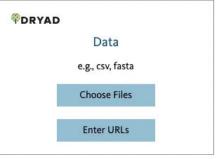
### **Upload Your Files**

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 <u>Zenodo</u>. You will have the opportunity to choose a separate license for your software on the review page.

- CCo license files; CCBY for Zenodo
- Open, acces types (non-proprie
- Tabular data validator automatically the format ar structure of 0 XLSX file format
- 300GB limit,







#### Files

Filename	Status	Tabular Data Check	URL	Туре	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				Data	672.26 KB	Remove

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

Upload pending files

## Review and next steps

Private for Peer Review option

#### Related Works

Article: https://doi.org/10.1016/j.mex.2022.10164

Edit Description

#### Review files



Data files hosted by Dryad

Fortuny-et-al-2021 Tomato-fruit-quality-traits-and-metabolite-content-are-affected-by-reciprocal-effect-and-heterosis RawData.xls 3.32 MB README.txt 2.07 kB

■ Edit Files

### Private for peer review

By choosing this option, your dataset will be private during your related article's peer review period. You will have access to a private dataset download URL to be shared with collaborators or the journal. Your dataset will not enter curation or be published. Because we may not have the status of your related article, the default for this period is six months. Please email us or uncheck this box at any point if your dataset is ready to enter curation.

☐ Keep my dataset private while my related article is in peer review

### Agree to terms

License and terms of service for data

☑ By checking this box, I agree to the license <u>CCO 1.0 Universal (CCO 1.0) Public Domain Dedication</u> \*

1 PUBLIC DOMAIN

☑ By checking this box, I agree to <u>Dryad's Terms of Service</u> \*

### Payment

Dryad charges a fee for data publication that covers curation and preservation of published datasets. Upon publication of your dataset, you will receive an invoice for \$120 USD. We're sensitive to the fact that fees for individual researchers can be a burden and create inequities. If you'd like to request a fee waiver, please contact help@datadryad.org.

- I agree to Dryad's payment terms
- Back to Upload

Submit

We simulated data with mean lead-time of 2.4 years and compared to observed data in the Nordic screening programs. They did not fit. First, we did not find any prevalence peak for women under age 60 when screening started in Sweden (1986-89), Norway (1996-99) and Denmark (2007-10), as opposed to the old RCTs from the 1970s and 1980s. The prevalence peaks for women over age 60 were much lower than in the RCts. In the old RCTs, most mammograpy detected tumors were also clinically detectable, and detection rates of DCIS were low. Second, in modern Nordic, public screening, detection rates are 2-3 times higher than in the old RCTs.

We suggest that the prevalence peaks in old days were due to the detection of many slow-growing clinical tumors with long lead times, and not early detection of subclinical tumors. Modern mammography detects a large number of tumors which never grow to become clinical cancers generating higher level of overdiagnosis than in the RCTs.

We calculated mean lead time in modern mammography screening. This has not been done before.

We also studied sensitivity using 2 different methods: We studie if interval cancer rates increased over times. We studied if many tumors slippped through one screening round because of low sensitivity abd were detected in the next screening round. Low sensitivity was not a major problem. We comcluded that low sensitivity is olny related to interval cancer and not to tumors wioth long lead time.

### Reviewer URL

You can share this version of your dataset files with others using the url below.

Click the Select Text button and then copy and paste the URL.

### Sharing Link

https://datadryad.org/stash/share/ozFdmYaevdRIhQBwVHdn-3L4tQoaCLD8f0s-RPKZaIU

Select Text





Search

[Daniella Lowenberg@dryad] | Admin ▼ |

Explore Data

About V

My Datasets

### Association between Procalcitonin Levels and Duration of Mechanical Ventilation in COVID-19 Patients

We are assembling your requested download for this dataset that is currently private for peer review.

If the download process does not begin automatically within several seconds, click the link above to start the download.



These additional files may be downloaded from Zenodo

dataset-analysis-output reports.Rmd



## Curation

Ensures data is appropriate and licensed for sharing

Verifies data is accessible and usable

Supports authors





## "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



## Publishing

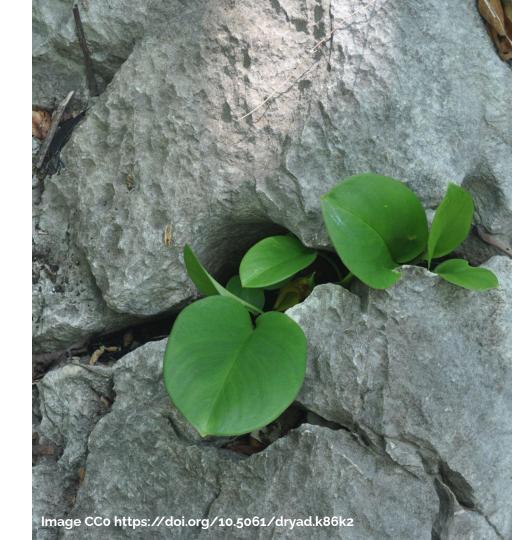
Ensures metadata quality
Increases discoverability
Connects data with other

research outputs

Promotes data citation

Makes data count







### ASHRAE global database of thermal comfort field measurements

Parkinson, Thomas, University of Sydney, bttps://orcid.org/0000-0002-0088-8754

Tartarini Enderica Parkalay Education Alliance for Decearch in Cingapara





## Characterizing and through microRNA

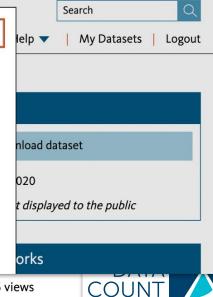
Nanayakkara, Jina, Queen's Unive Yang, Xiaojing, Queen's University Tyryshkin, Kathrin, Queen's Unive Wong, Justin J.M., Queen's Univer Vanderbeck, Kaitlin, Queen's Univ Ginter, Paula S., Weill Cornell Med Scognamiglio, Theresa, Weill Corn

### **Works Referencing This Dataset**

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

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

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, <a href="https://doi.org/10.1002/gcc.22189">https://doi.org/10.1002/gcc.22189</a>





### Citation

Parkinson, Thomas et al. (2022), ASHRAE global database of thermal comfort field measurements, Dryad, Dataset, https://doi.org/10.6078/D1F671





similarities. Using machine learning approaches, we identified 17 miRNAs to discriminate 15 NEN pathological types and subsequently constructed a multi-layer classifier, correctly identifying 217 (98%) of 221 samples and overturning one histologic diagnosis. Through our research, we have identified common and type-specific miRNA tissue markers and constructed an accurate miRNA-based classifier, advancing our understanding of NEN diversity.

#### Methods

Sequencing-based miRNA expression profiles from 378 clinical samples, comprising 239 neuroendocrine neoplasm (NEN) cases and 139 site-matched non-NEN controls, were used in this study. Expression profiles were either compiled from published studies (n=149) or generated through small RNA sequencing (n=229). Prior to sequencing, total RNA was isolated from formalin-fixed paraffin-embedded (FFPE) tissue blocks or fresh-frozen (FF) tissue samples. Small RNA cDNA libraries were sequenced on HiSeq 2500 Illumina platforms using an established small RNA sequencing (Hafner et al., 2012 *Methods*) and sequence annotation pipeline (Brown et al., 2013 *Front Genet*) to generate miRNA expression profiles. Scaling our existing approach to miRNA-based NEN classification (Panarelli et al., 2019 *Endocr Relat Cancer*, Ren et al., 2017 *Oncotarget*), we constructed and cross-validated a multi-layer classifier for discriminating NEN pathological types based on selected miRNAs.

### **Usage Notes**

Diagnostic histopathology and small RNA cDNA library preparation information for all samples are presented in Table S1 of the associated manuscript.

### **Funding**

Academic Health Sciences Center Alternative Funding Plan Innovation Fund

Canada Foundation for Innovation John R Evans Leaders Fund

Carcinoid and Neuroendocrine Tumor Society Canada

Ontario Research Fund-Research Infrastructure

Southeastern Ontario Academic Medical Organization

Robertson Therapeutic Development\*

Rockefeller University, Award: UL1TR001866

Ontario Institute for Cancer Research

Robertson Therapeutic Development

#### License

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







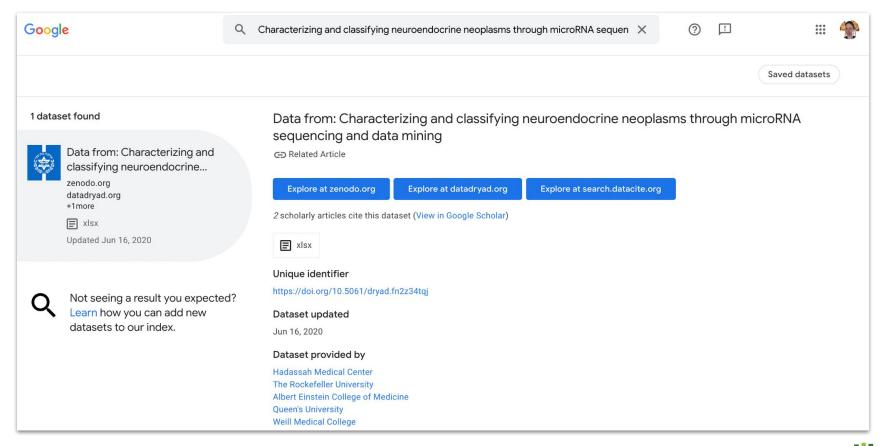
## Reach

Indexed by relevant aggregators

Discoverable in Google, Google Dataset Search, Scopus, Web of Science, and more









### Admin Dashboard

At a glance

82,796 Users 71,005 Datasets Activity in the last 7 days

244 users added 209 datasets started 128 datasets submitted

### **Datasets**

Search Terms:		terms †	Search	Clear Search		
Filter by: Institution	~	Curator:	Curator	~	Status: Curation	~
Journal Name:						

### Reset all filters

<u>Title</u>	Status	Author	DOI	Last Modified	Curator	Size	Publication Date
Why flowers close at noon? A case study of an alpine species Gentianopsis paludosa (Gentianaceae)	Curation 🌶	Duan; Ehmet; Hou; Pang; Shao; Sun; Zhao	10.5061/dryad.qv9s4mwg4	03/08/2022 11:31:06 UTC	Jessica Herzog	615.30 kB	January 15, 2022
CV and colony data vespid mandible wear	Curation 🖋	Lagos-Oviedo; Sarmiento	10.5061/dryad.ffbg79csk	03/13/2022 23:32:04 UTC O	Jessica Herzog	293.48 kB	Not available
Seismic Data Imaging the Earth: PmPWorld	Curation &	Yang	10.25349/D9QK7K	03/21/2022 16:52:05 UTC	Ryan Horne	12.45 MB	Not available

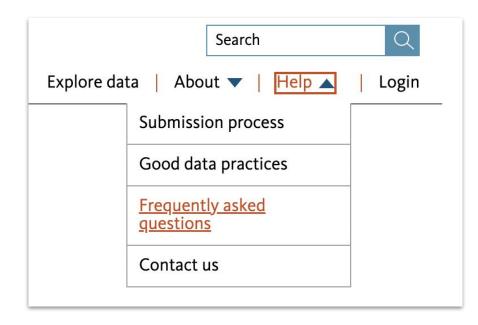


# We're here to help

Help & support <u>help@datadryad.org</u>

General inquiries <a href="hello@datadryad.org">hello@datadryad.org</a>

Community <u>sarah@datadryad.org</u>





## ARCHIVING DATA

Prof. Andrew Beckerman
School of Biosciences
Board of Directors, DRYAD
Editor in Chief, Ecology and Evolution
(Wiley - British Ecol Sco - Eur Soc Evol Bio)

# ARCHIVING IS HARD WORK









# LUCKY!

#### **Publisher members**

American Academy for the Advancement of Science

American Genetic Association

American Geophysical Union

American Medical Informatics

Association

American Ornithological Society

American Society of Naturalists

American Society of Plant

**Taxonomists** 

Association for Tropical Biology and Conservation

**British Ecological Society** 

Cambridge University Press

eLife

**European Society for Evolutionary** 

Biology

Hindawi

Institute of Botany, Chinese

Academy of Sciences

International Society for Behavioral Ecology

John Wiley & Sons, Inc

**Nordic Society Oikos** 

Oxford University Press

The Paleontological Association

The Paleontological Society

**Rockefeller University Press** 

The Royal Society

Society for the Study of Evolution

Society of Systematic Biologists

The University of Chicago Press

# ARCHIVING IS HARD WORK

# WORK HARDER...

- Tidy Data in Findable, Accessible, Interoperable, and Re-usable format (FAIR).
  - NOT PDF, not xlsx, not platform specific format.
- Find a place to store it
  - The repository
- Possibly pay for the honour, service and benefits.

# WORK HARDER...

- Tidy Code/Photos/Pipeline in (FAIR) format:
  - Findable, Accessible,
     Interoperable, and
     Re-usable.
  - NOT PDF, not xlsx, not platform specific.
- Find a place to store it
  - The repository
- Possibly pay for the honour, service and benefits.

### OLD DEFINITION

The Methods Section.

- Sufficient description
   of the processes by
   which the figures and
   inference in your
   manuscript were
   developed.
- Sufficient = SOMEONE ELSE CAN MAKE IT HAPPEN.
- Maybe somebody got in touch to collaborate or ask for the data

## OLD DEFINITION

The Methods Section.

- Sufficient description of the processes by which the figures and inference in your manuscript were developed.
- Maybe? Definitely
   Hopefully someone can use your data...
- DATA ARCHIVING IS VALUABLE.
  - $\circ$  Making the data
  - Translating the data into figures, graphs, tables and inference.

# PERCEIVED RISKS AND BENEFITS

## RISKS

Somebody does something with your data before you do.

Somebody finds errors.

Somebody makes money and you don't.

### HAVE YOU EVER USED SOMEONE ELSE'S DATA?

DO YOU KNOW ANYONE WHO HAS USED SOMEONE ELSE'S DATA?

DO YOU KNOW ANYONE WHO'S DATA HAS BEEN USED?

## RISKS

- Somebody does something with your data before you do.
  - EMBARGOS
  - ATTRIBUTION
- Somebody finds errors.
  - Better late than never
  - Improved science

# BENEFITS

- Somebody does something with your data
- Somebody finds errors.

### BENEFITS

- Somebody does something with your data
  - ∘ Meta-analysis
  - New analysis
  - People without good libraries can benefit
    - Global South researchers
    - Policy/ Government/ Business
  - Collaboration
  - Attribution
- Somebody finds errors.
  - Better late than never
  - Improved science

# THE OPEN DATA (RESEARCH) LANDSCAPE

# WHO'S INVOLVED AND WHY ALL THE HOO-HA?

Governments

Gov and Non-Gov Funders

Publishers

Societies

Data and Code Repositories

Researchers

Funders
(UKRI, BBSRC, NERC,
EPSRC, Royal Society,
EU)

Societies (Genetics, SEB, BES, ESEB, EGU)

Publishers (Wiley, OUP, Elsevier, PLOS)

Advocates (OSF, rOpenSci)

Data Repositories and Data Publishers, Code Publishers (DRYAD, NCBI, Zenodo, FigShare, GitHub)

Journals Society + Wiley, OUP, Elsevier

Researchers (You!)

# THE BURDEN IS ON YOU/US AT THE MOMENT... THIS WILL CHANGE



#### About Plan S

Plan S is an initiative for Open Access publishing that was launched in September 2018. The plan is supported by cOAltition S, an international consortium of research funding and performing organisations. Plan S requires that, from 2021, scientific publications that result from research funded by public grants must be published in compliant Open Access journals or platforms.

Read more

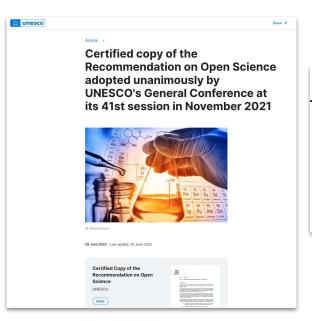


#### Policy

### **UKRI** open access policy

UKRI's open access policy for research publications that acknowledge funding from UKRI and any of its councils.

From: UKRI
Published: 6 August 2021





# WHO'S INVOLVED AND WHY ALL THE HOO-HA?

Pay Attention, Get
Involved, Know the Options

Talk to the Library!

Talk to your primary academic society and publishing team.

Have a lab meeting about OA, OD, OR.



# Specifying a repository in your DMP

Scholarly Communications Team
The University Library



### Repository choice in Data Management Planning

Factors for your DMP

- Subject is there an appropriate repository?
- Sensitivity does access need to be restricted?
- Collaborators where are they sharing their data?
- Selection for archiving can all data be shared in the same way?

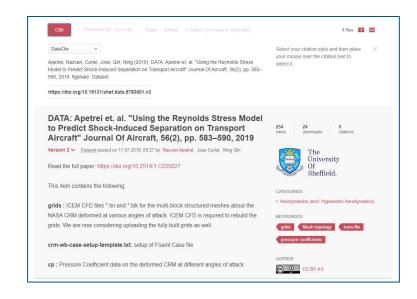
Remember ...

- You may not need to choose a repository, but you should describe it
- You may need more than one solution
- Your choices may affect the metadata that you collect



### Depositing data in ORDA - orda.shef.ac.uk

- Creates a DOI for your data
- Offers options for
  - level of public access
  - licence e.g. Creative Commons
  - embargo period
- Record details in ORDA, even if data is stored elsewhere
- More information on the RDM webpages





### We're here to help

#### **Email**

rdm@sheffield.ac.uk

#### Website

sheffield.ac.uk/library/research-data-management

### **DMPs**

sheffield.ac.uk/library/research-data-management/planning

### **Training**

sheffield.ac.uk/library/research-data-management/training

