

Data Sharing: Best Practices and Measuring Impact

September 7th 2023

Agenda

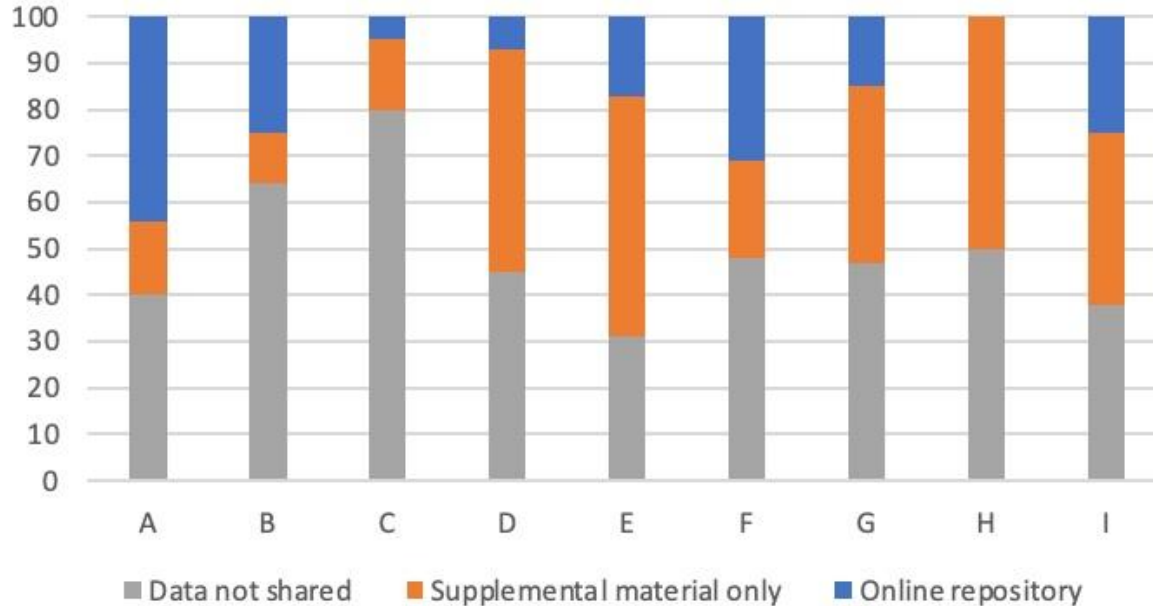
- **Introduction and Data Sharing BOF report** - Kristen Ratan, Stratos
- **Data sharing workflow overview** - Jennifer Kemp, Stratos
- **Dryad** - Jess Herzog, Dryad
- **DataCite** - Xiaoli Chen, DataCite
- **Make Data Count** - Iratxe Puebla, Make Data Count Initiative
- **Wrap up** - Kristen Ratan, Stratos

Spring: Data Sharing Birds of a Feather (BOF)

- 9 HRA members participated
- Three main topics covered
 - Policy-making - What to prioritize, consequences of policy choices, how to create and communicate policies
 - Best Practices and Implementation - Grantee guidance and support, Infrastructures, tools, and partners, post-grant period
 - Compliance and Impact Tracking - How to measure data sharing, compliance, what tool chain to use, what are the metrics
- Actions:
 - Updated the [Policy Worksheet](#) to include OSTP and NIH info
 - Cohort analysis of current compliance with Nelson, open access and data sharing

DataSeer Report on BOF Cohort Data Sharing

Percent Articles Sharing data, by Funder



- 99% generated data
- 60% overall shared data

Of these:

- 30% of authors shared their data only as supplemental files
- 3% shared some portion of their data as supplemental files and shared other datasets on an online repository.
- 6% of authors put all of their dataset in a repository

And

- 41% shared no data at all

Coming Next:

Fall Data Sharing BOF

Data sharing workflow

Funding

- Preliminary data
- Data sharing plan

Research

- Dataset creation & collection
- Data storage

Analysis & Writing

- Preparing to share findings
- Data curation

Publishing/ Sharing

- Dataset in a FAIR repository
- Data citation

Post-Publication

- Data citation
- Re/use metrics



Funders,
Grantees

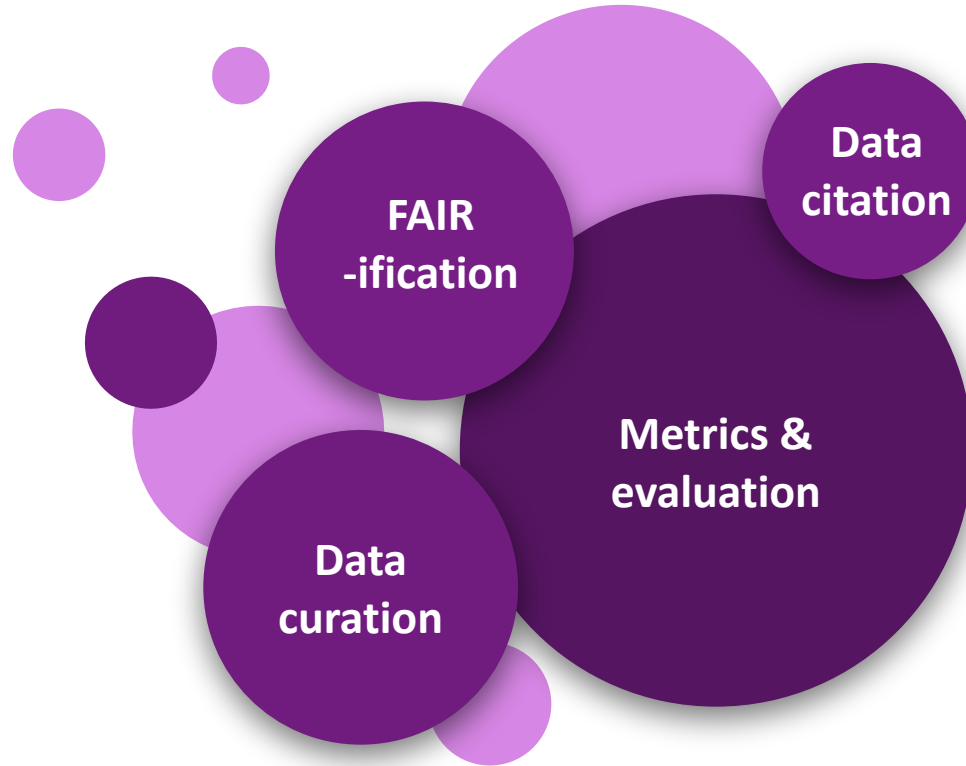
Researchers,
Institutions

Co-/Authors,
Data helpers

Publishers,
Infrastructures

Many
stakeholders

For discussion: What are the needs/gaps?



A little bit about Dryad

- Open research data platform
- Community-driven effort to make research data discoverable, freely reusable, and citable
- Non-profit committed to the values of open science: infrastructure and research data is open source and open access
- Data is curated, protected, and perpetually preserved
- Published datasets with no barriers for reuse (CC0 license waiver)

+46k datasets have been curated and published by Dryad

Datasets deposited by **+193k researchers** worldwide

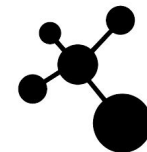
Submissions from **+69k international institutions**

Datasets associated with **+1,200 academic journals**



Dryad's mission is to enable and promote the re-use of research data

Dryad creates trusted connections to research data



ASHRAE global database of thermal comfort field measurements

Parkinson, Thomas, University of Sydney, <https://orcid.org/0000-0002-0088-8754>
Tartarini, Federico, Berkeley Education Alliance for Research in Singapore
Földváry Ličina, Veronika, University of California, Berkeley
Cheung, Toby, Dayton Foundation, <https://orcid.org/0000-0003-0756-0214>
Zhang, Hui, University of California, Berkeley
de Dear, Richard, University of Sydney, <https://orcid.org/0000-0002-3414-290X>
Li, Peixian, University of British Columbia, <https://orcid.org/0000-0002-5473-1605>
Arens, Edward, University of California, Berkeley
Chun, Chungyoon, Yonsei University
Schiavon, Stefano, University of California, Berkeley, <https://orcid.org/0000-0003-1285-5682>
Luo, Maohui, University of California, Berkeley
Brager, Gail, University of California, Berkeley, <https://orcid.org/0000-0002-1100-8302>
tom.parkinson@berkeley.edu, federico.tartarini@bears-berkeley.sg, veronika.foldvary@gmail.com,
toby.cheung@bears-berkeley.sg, zhanghui@berkeley.edu, richard.dedear@sydney.edu.au,
ariel.li@berkeley.edu, earens@berkeley.edu, chun@yonsei.ac.kr, schiavon@berkeley.edu,
lmhtongji@berkeley.edu, gbrager@berkeley.edu

Publication date: July 15, 2022

Publisher: Dryad

<https://doi.org/10.6078/D1F671>



Citation

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

Data files

Download dataset

- > July 1, 2018
- > July 1, 2018
- > July 1, 2018
- > July 1, 2018
- > June 9, 2022

ORCID

ROR

Related Works

Article
<https://doi.org/10.10...buildenv.2018.06.022>

Software
<https://github.com/Ce...ronment/ashrae-db-ii>

Connecting related works like: DMPs, preprints, supplemental info, other data, etc.

Metrics

9916 views

2108 downloads

3 citations



DC¹
Data Citation Principles



Data — connected



- Dryad requires the submitting author to log in using their **ORCID iD** and invites co-authors to also attach their iDs
- Curators match institutional affiliations to **ROR IDs**
- Curators connect funder info to Crossref's Funder Registry (**FundRef**)
- Published datasets are indexed by search engines including **Scopus, the Thomson-Reuters Data Citation Index, and Google Dataset Search**
- Related works can be **linked** from the dataset landing page to the article DOI, Data Management Plans (DMPs), preprints, software, scripts, code, supplementary material and other related research objects

Dryad curators

Idea-generators, skilled communicators, natural helpers, proactive, dedicated, diverse, adaptable, supportive, and motivated team members



Experience and/or advanced degrees in:

- Information Technology
- Teaching/training
- Chemistry
- Biochemistry
- Geology
- Ecology & Evolutionary Biology
- Accounting
- MLIS
- Civil & Environmental Engineering

Ensure data is organized, formatted, appropriate, and properly licensed for sharing

Verify that data files are accessible, explained, and usable

Offer personalized, knowledgeable support and guidance to authors

Link metadata to trusted PIDs to promote discoverability and connectivity

Evaluate submissions beyond a checklist
Identify and report issues for investigation;
suggest improvements to optimize process for authors and curation team

Human subjects data

Human subjects data must be properly anonymized and prepared under applicable legal and ethical guidelines in order to publish in Dryad

This can be challenging in scenarios when identifiers included in a dataset are essential to data reuse

Curators work with authors to ensure their data meet Dryad's HSI guidelines — with the goal of maximizing reusability without reducing its intended value to the research community

AUTHOR RESOURCES



Best Practices: Sharing Human Subjects Data

Dryad Requirements

Researchers are responsible for ensuring that all contents of their data package do not contain information that can be used alone, or in aggregate, to identify any individual.

Dryad's policies on human subjects data are in accordance with accepted international standards for de-identifying data from such trusted sources as [General Data Protection Regulation \(GDPR\)](#), [HIPAA privacy rules](#), the [Act on the Protection of Personal Information \(APPI\)](#), and the [Personal Information Protection Law \(PIPL\)](#). Dryad will uphold the policies and publication requirements set in keeping with our responsibility to protect human participants and maintain the integrity of the research data we publish – regardless of whether the data submitted is or will be openly available elsewhere.

Preparing Your Data

Dryad does not publish any direct identifiers. A direct identifier is information that is sufficient, on its own, to disclose the identity of a research participant. Examples include: name, address, postal code, telephone number, voice, video, or photograph.

Datasets may contain **no more than three indirect identifiers**, such as demographic, biological, and geographic data, that could lead to re-identification if combined with other available data (either collected as part of your research, or available elsewhere). Examples include: institutional affiliations, occupation, geographic region, unique values or characteristics (outliers).

To properly de-identify your data, consider direct and indirect identifiers and evaluate whether the combination of identifiers could lead to re-identification. For example, the age of participants, uncommon characteristics of the individual (e.g., rare health condition, number of children), geographic/regional location, named facility and/or service provider, and highly visible characteristics of the individual (e.g., ethnicity, race).

A partial listing of common **direct and indirect identifiers** is provided in the table on [pages 2-3](#). Because this is not a definitive list of potentially concerning identifiers, we recommend referencing the table to guide your understanding of the type of variables that can be identifying and help you recognize and categorize other direct or indirect identifiers in your data.

A detailed description of your process for de-identifying data should be included in your README file. Additionally, if your research funder requires a [Data Management Plan \(DMP\)](#), statements of protections for privacy, rights, and confidentiality of human research participants will be required. Click [here](#) to view an example of a publicly available DMP.

CURATOR RESOURCES

Human Subjects Information (HSI) Dataset Curation Checklist

Last Updated: May 2023

Note: This is the master template. To create a working copy for your own use, click 'File', then click 'Make a copy' and add the DOI of the submission to the title of your version to avoid confusion.

Also, additional guidance is provided by the Data Curation Network (DCN) [here](#).

Review metadata for HSI-specific information, paying close attention to the author affiliations, abstract and methods, which will help to contextualize the riskiness of the data:

- Does the metadata describe the research population and their inclusion criteria (usually in the abstract or methods)?
- Were participants recruited from a vulnerable population (children, adolescents, the elderly, imprisoned people, patients with a rare condition, etc.)?
- Do the author affiliations reveal the recruitment pool for the subject population?

Human Subjects Information (HSI) Dataset Curation Checklist

Last Updated: May 2023

Check data for

- Are the _____ should be aggregated or made less specific.
- Are there any variables that are not identifiers in isolation, but could be identifying when combined with other variables? Count these as indirect identifiers.

Check any supplemental or software (Zenodo files) for [direct identifiers](#) (none allowed) and [indirect identifiers](#) (3 allowed):

- Are there any direct identifiers? If so, what are they?

- Are there any indirect identifiers? If so, what are they?

Check data for

- Are the _____
- Is the _____
- Are the _____

README check (README is required for HSI datasets):

- Does the README clearly define all variables and values included in the data?
- Does the README reveal additional information about the subject population not included elsewhere?

(next page)

Ask yourself, with the information provided in this dataset, could I...

- Potentially identify an acquaintance, like a co-worker or neighbor?
- Potentially identify any of the participants by looking up their information online?

If you've answered 'yes' to either, this dataset has not been sufficiently de-identified.

Based on your review, which edits are needed from the author(s)?

- Direct identifiers need to be removed.
- Indirect identifiers need to be reduced to 3 or fewer.
- Variables must be aggregated (like grouping ages) or made less precise.
- Providing a comprehensive README or improving an existing README.
- Providing a blank copy of the study consent form to review terms of consent - have the participants consented to their data being published? For guidance on consent forms, visit [this link](#).
 - When do you request a consent form? When there is a vulnerable population involved and/or if you have any doubt concerning whether the participants provided consent.
 - If you are unable to confirm whether the author obtained consent, in



Dryad for funders

Capturing metadata: funding



Funding

Granting organization: *	Award number	Program/division
<input type="text" value="National Institutes of Health"/>	<input type="text" value="3OT2DB000005-01S2"/>	<input type="text"/>
NIH Institute or Center		
<input type="text" value="- Select one -"/>		
remove		

No funding received

Depending on the funder information input into the “granting organization” field, a second drop-down menu will appear to collect more precise information.

If no funding was secured, a checkbox is available to bypass the required metadata field.

Capturing metadata: funding



Citation

Cao, Shutian et al. (2023), Data from: Effects of temperature and obsidian content on the friction and stability of simulated basalt gouges: Implications for shallow moonquakes, Dryad, Dataset, <https://doi.org/10.5061/dryad.bvq83bkfn>

Abstract

Basalt is a major component of crust on both the Earth and Moon. Mineral composition and temperature influence frictional instability and thus the potential for seismicity on basaltic faults. We performed velocity-stepping shear experiments on basalt gouges at a confining pressure of 100 MPa, temperatures in the range 100-400°C and with varied obsidian contents of 0-100 wt.% under wet/dry conditions to investigate the frictional strength and stability of basaltic faults. We observe a transition from velocity-neutral to velocity-weakening behaviors with increasing obsidian content. The frictional stability response of the mixed obsidian/basalt gouges is characterized by a transition from velocity-strengthening to velocity-weakening at 200°C and another transition to velocity-strengthening at >300°C. Conversely, frictional strengths of the obsidian-bearing gouges are insensitive to temperature and wet/dry conditions. These results suggest that obsidian content dominates the potential seismic response of basaltic faults with the effect of temperature controlling the range of seismogenic depths. These observations contribute to a better understanding of the nucleation mechanism of shallow moonquakes and also seismicity on terrestrial faults in basalt.

Funding

National Natural Science Foundation of China, Award: 42077247

Fundamental Research Funds for the Central Universities

G. Albert Shoemaker endowment*

National Natural Science Foundation of China, Award: 42107163

Subject keywords

Earth and related environmental sciences
basaltic faults
Frictional stability
Mineral components
shallow moonquakes
temperature effect

License

This work is licensed under a [CC0 1.0 Universal \(CC0 1.0\) Public Domain Dedication](https://creativecommons.org/licenses/by/4.0/) license.



Published datasets will display funders and award numbers; unique entries not matched to ROR are denoted by an asterisk

Capturing metadata: funding



```
"fundingReferences": [{
  "schemeUri": "https://www.crossref.org/services/funder-registry/ ",
  "funderName": "National Natural Science Foundation of China" ,
  "awardNumber": "42077247",
  "funderIdentifier": "https://doi.org/10.13039/501100001809 ",
  "funderIdentifierType": "Crossref Funder ID"
},{
  "schemeUri": "https://www.crossref.org/services/funder-registry/ ",
  "funderName": "Fundamental Research Funds for the Central
Universities",
  "funderIdentifier": "https://doi.org/10.13039/501100012226 ",
  "funderIdentifierType": "Crossref Funder ID"
},{
  "funderName": "G. Albert Shoemaker endowment"
},{
  "schemeUri": "https://www.crossref.org/services/funder-registry/ ",
  "funderName": "National Natural Science Foundation of China" ,
  "awardNumber": "42107163",
  "funderIdentifier": "https://doi.org/10.13039/501100001809 ",
  "funderIdentifierType": "Crossref Funder ID"
}
]
```



new!

Funder facet: via Explore data



Explore data

Search



Who we are

What we do

Join us

Help

Datasets

Logout

National Institutes of Health

Search

Start Over

National Institutes of Health



Limit your search

< Previous | 1 - 10 of 1,285 | Next >

Subject keyword



Geographical Location



Journal



Institution



File Extension



Funder



National Institutes of Health

1,165

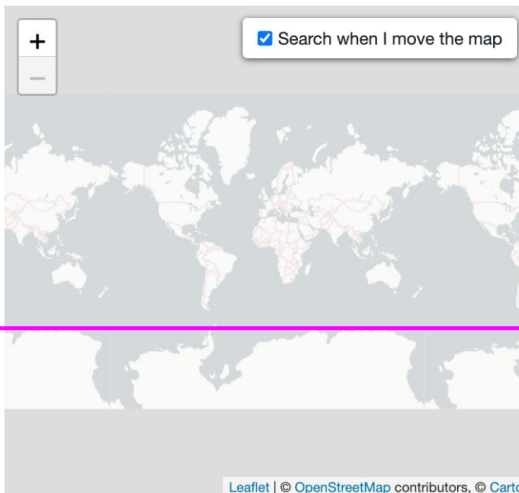
National Science Foundation

319

National Institute of General Medical Sciences

146

1. Propagating spatio-temporal activity patterns across macaque motor cortex carry kinematic information
2. SAFER: Safer conception strategies to prevent HIV transmission among HIV sero-different couples desiring pregnancy
3. Data from: Implementation of a pediatric telemedicine and medication delivery service in a resource-limited setting: A pilot study for clinical safety and feasibility
4. Oral cholestyramine prevents the enrichment of diverse daptomycin-resistance mutations in intestinal *Enterococcus faecium* populations
5. Arginine-vasopressin expressing neurons in the murine suprachiasmatic nucleus exhibit a circadian rhythm in network coherence in vivo
6. Incucyte data for progeny virus derived from insect cells with or without methyltransferase inhibitor



Leaflet | © OpenStreetMap contributors, © Carto

Introduced new option to **search** for and **filter** on **Funder**



Funder dashboard: search & find



DRYAD Explore data

Who we are | What we do | Join us | Help | Datasets | Logout

Dataset funder dashboard

Limit to

Partner institution:

Funder:

For dates

Date type:

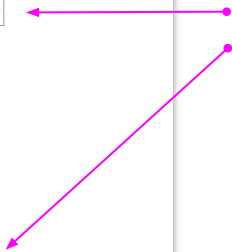
Start date:

to

End date:

Title ↑↓	Authors ↑↓	DOI ↑↓	Funder ↑↓	Award ↑↓	Submitted ↑↓	Embargoed ↑↓	Published ↑↓
β-cell-specific deletion of Zfp148 improves nutrient-stimulated β-cell Ca2+ responses	Attie; Cardone; de Klerk; Emfinger; Foster; Gygi; Hebrok; Keller; Kendziorski; Kibbey; Lewandowski; Liu; Merrins; Mitok; Paulo; Perales; Rabaglia; Schueler; Simonett; Stapleton; Wang; Yu	10.5061/dryad.bcc2fqzcv	National Institute of Diabetes and Digestive and Kidney Diseases	R01DK101573	08/16/2021		05/11/2022
β-cell-specific deletion of Zfp148	Attie; Cardone; de Klerk; Emfinger; Foster; Gygi; Hebrok;	10.5061/dryad.bcc2fqzcv	National Institute of	R01DK127637	08/16/2021		05/11/2022

Options to focus search and sort on columns





Funder dashboard: export data



Search results can be exported in CSV format

	A	B	C	D	E	F	G	H	I
1	Title	Authors	DOI	Funder	Funder id	Award	First submitted	Embargoed	Published
2	Data from: Age-at-injury	Green; Murphy; Ortiz; Ro	10.5061/dryad.5tb2rbp4r	National Institute of Neu	http://dx.doi.org/10.130: R21NS120022		08/25/2021 06:01:03 UTC		01/01/2022 00:00:00 UTC
3	Data from: Use of an exc	Bowers; Carper; Muchoni	10.5061/dryad.7pvmcvdx	National Institute of Gen	http://dx.doi.org/10.130: GM103440		12/26/2022 20:25:59 UTC		01/01/2023 00:00:00 UTC
4	Quantitative analysis of	Leirit; Ryder	10.5061/dryad.h70rxwdg	National Institute of Gen	http://dx.doi.org/10.130: 5K12GM000680		07/06/2020 21:30:03 UTC		01/02/2021 00:00:00 UTC
5	Quantitative analysis of	Leirit; Ryder	10.5061/dryad.h70rxwdg	National Institute of Gen	http://dx.doi.org/10.130: 1F32GM128407		07/06/2020 21:30:03 UTC		01/02/2021 00:00:00 UTC
6	Quantitative analysis of	Leirit; Ryder	10.5061/dryad.h70rxwdg	National Heart, Lung, and	http://dx.doi.org/10.130: 5K22HL126922		07/06/2020 21:30:03 UTC		01/02/2021 00:00:00 UTC
7	Trans-specific polymorph	Lively; Million	10.5061/dryad.t76h8r2f	Eunice Kennedy Shriver N	http://dx.doi.org/10.130: T32 HD049336		12/09/2021 01:41:04 UTC		01/02/2023 00:00:00 UTC
8	Data from: Total syntheses	Adamson; Darzi; Donalds	10.5068/D1Q38N	National Institute of Gen	http://dx.doi.org/10.13039/100000057		01/19/2022 00:32:04 UTC		01/03/2023 00:00:00 UTC
9	Summary statistics from	Ambati; Faraco; Hallmay	10.5061/dryad.kd51c5b9l	National Institutes of Hei	http://dx.doi.org/10.130: NIH-23724		10/29/2022 01:06:04 UTC		01/03/2023 00:00:00 UTC
10	Summary statistics from	Ambati; Faraco; Hallmay	10.5061/dryad.kd51c5b9l	National Institute of Mer	http://dx.doi.org/10.130: 58C2MH089916		10/29/2022 01:06:04 UTC		01/03/2023 00:00:00 UTC

- Title
- Authors
- DOI
- Funder
- Funder iD
- Award
- Date submitted
- Date embargoed
- Date published

Dryad funder-focused use cases



As a funder from a specific NIH institute or in general, I want to find datasets we have funded in Dryad, so that I can **REPORT** on compliance with policies, and **TRACK** impact of research funding and usage of data.

This use case highlights ways funders can leverage generalist repositories to track compliance with data sharing policies and understand data reuse.

Use case date:
May 2023

Use case contact:
hello@datadryad.org

Background:
As a funding body, I want to track compliance with my open data sharing policy and understand more about how and where researchers receiving funding are sharing their data.

Use Case:
Dryad integrates with Crossref's Funder Registry to provide standardized, machine-readable funding information about the datasets we publish. Funders can easily identify the data they sponsor by executing a blank search at <https://datadryad.org/search> (visit the link and click the magnifying glass icon without entering a search term) and then filtering by "Funder". Dryad members, including funders, can also gain access to our interactive administrative dashboard, which offers detailed information about datasets affiliated with the organization, including dataset metadata and usage metrics.



Screenshot of the "Funder" filter in Dryad's search interface, which allows easy identification of datasets funded by a wide range of organizations.

Funder	Count
National Institutes of Health #	1,077
National Science Foundation	275
National Institute of General Medical Sciences	125



GREI Use Cases are supported by the National Institutes of Health (NIH) Office of Data Science Strategy/ Office of the NIH Director pursuant to OTA-21-009, "Generalist Repository Ecosystem Initiative (GREI)". NIH Image Gallery: GFP10 protein forms new brain cell connections <https://www.nih.gov/photos-images/2024/07/29/20240729-01>. Courtesy of S. Thomas Carmichael, MD, PhD, David Geffen School of Medicine at the University of California Los Angeles. NIH Funding: National Institute of Neurological Disorders and Stroke (NINDS). More information at <https://www.nih.gov/news-events/news-releases/scientists-identify-main-component-brain-repair-after-stroke>



As an NIH-funded researcher, I want to **USE** Dryad to share my data, so that I can comply with my data management and sharing plan and the conditions of my grant.

This use case highlights ways research teams can leverage generalist repositories to share project data.

Dataset title:
Deep evolutionary analysis reveals the design principles of fold A glycosyltransferases <https://doi.org/10.5061/dryad.v15dv41sh>

Investigators and affiliations:
Taujale, Rahil et al.

Publication date:
April 10, 2020

Data type:
FASTA and TSV

Citations:
Rahil Taujale, et al. (2020). "Deep evolutionary analysis reveals the design principles of fold A glycosyltransferases," *eLife* <https://doi.org/10.7554/eLife.54532>

Use case date:
April 2023

Use case contact:
hello@datadryad.org

Background:
As an NIH-funded researcher, I am required to submit my data to an approved repository. I have FASTA and TSV files containing GT-A fold sequences and have determined that a suitable domain-specific repository does not exist. I have confirmed that the data is shareable under a CC0 license and does not contain any sensitive or personally identifying information. I want to ensure that my dataset is discoverable, citable, and connected to the research article it underlies.

Process:
I login to datadryad.org using my ORCID ID and check to see if my institution is a Dryad member (https://datadryad.org/stash/oin_us#members). I complete Dryad's straightforward submission process, which collects metadata necessary for discovery and reuse. I indicate NIH as a granting organization and enter my grant number. I upload my data files and a detailed README describing my methods, the file structure and contents of my submission, variable definitions, and other information necessary for reuse. Through Dryad's integration with Zenodo, I also seamlessly upload code needed to replicate my analyses. I receive a DOI upon submission.

Within a few days, a Dryad curator will confirm whether my data are appropriate for open sharing, follow FAIR principles, and meet ethical standards for publication. They will also offer guidance on best practices for creating reusable data and help me navigate publication requirements. They will not attempt to assess the rigor of my methods or validate my findings.

My data will be published openly via datadryad.org.

Login	Submit	Review	Cite
Use your ORCID. If your institution is a Dryad member , connect to your existing credentials.	Upload your data files and receive a citable DOI.	Our curators will thoroughly check your submission to ensure the data are appropriate and ready for sharing and reuse. They may contact you with advice or questions.	Cite and promote your data publication.



GREI Use Cases are supported by the National Institutes of Health (NIH) Office of Data Science Strategy/ Office of the NIH Director pursuant to OTA-21-009, "Generalist Repository Ecosystem Initiative (GREI)". NIH Image Gallery: GFP10 protein forms new brain cell connections <https://www.nih.gov/photos-images/2024/07/29/20240729-01>. Courtesy of S. Thomas Carmichael, MD, PhD, David Geffen School of Medicine at the University of California Los Angeles. NIH Funding: National Institute of Neurological Disorders and Stroke (NINDS). More information at <https://www.nih.gov/news-events/news-releases/scientists-identify-main-component-brain-repair-after-stroke>

Use cases designed for **NIH funders** and **NIH-funded researchers**

Hey, funders!



Tell us more about:

1. Your workflow process
2. Who the key administrative stakeholders are within your organization
3. What features or functions you'd like to see in a funder dashboard

Thanks for listening.

hello@datadryad.org



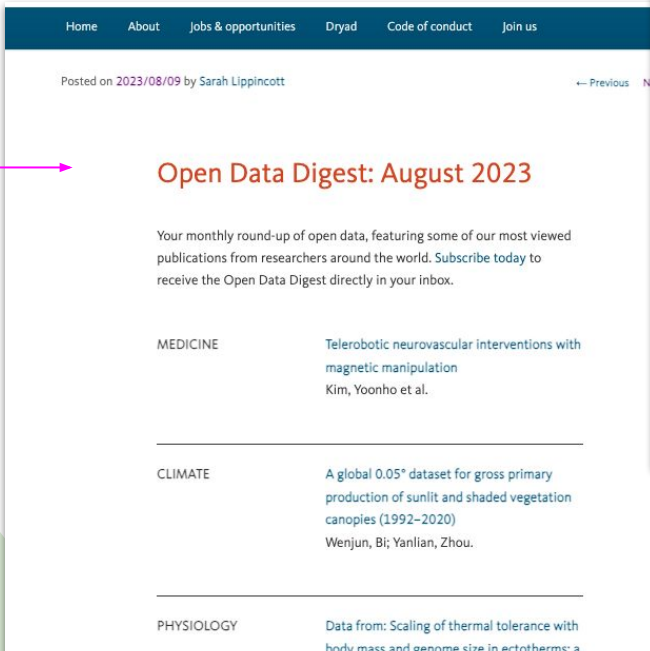
<https://blog.datadryad.org/about/subscribe/>

Don't miss an update

Subscribe to the Dryad newsletter and Open Data Digest to get compelling data publications, tips and resources, event invitations, and more delivered straight to your inbox.

Sign me up for the Open Data Digest

Sign me up for the Dryad Newsletter



Home About Jobs & opportunities Dryad Code of conduct Join us

Posted on 2023/08/09 by Sarah Lippincott

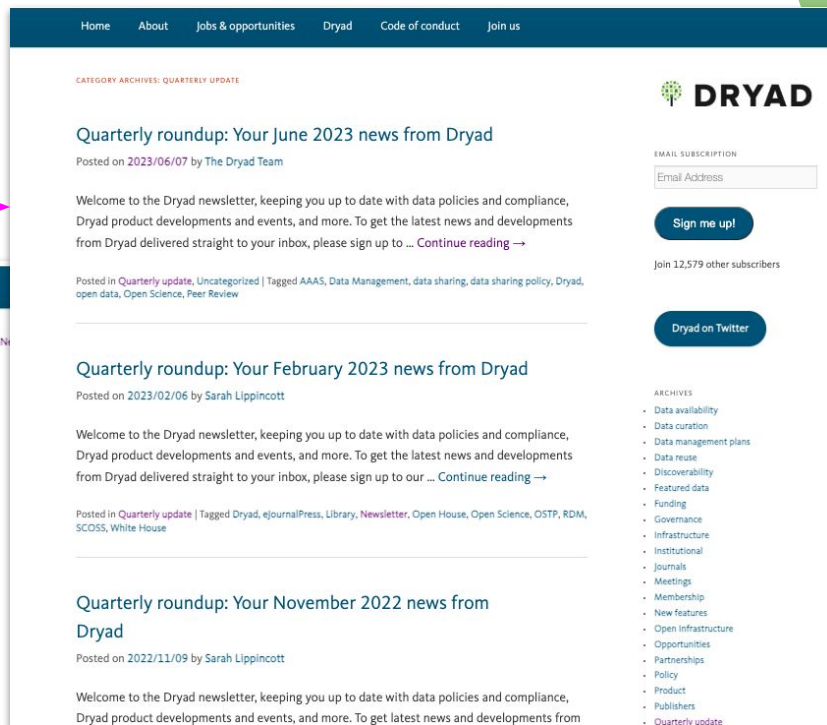
Open Data Digest: August 2023

Your monthly round-up of open data, featuring some of our most viewed publications from researchers around the world. Subscribe today to receive the Open Data Digest directly in your inbox.

MEDICINE Telerobotic neurovascular interventions with magnetic manipulation
Kim, Yoonho et al.

CLIMATE A global 0.05° dataset for gross primary production of sunlit and shaded vegetation canopies (1992–2020)
Wenjun, Bi; Yanlian, Zhou.

PHYSIOLOGY Data from: Scaling of thermal tolerance with body mass and genome size in ectotherms: a



Home About Jobs & opportunities Dryad Code of conduct Join us

CATEGORY ARCHIVES: QUARTERLY UPDATE

Quarterly roundup: Your June 2023 news from Dryad

Posted on 2023/06/07 by The Dryad Team

Welcome to the Dryad newsletter, keeping you up to date with data policies and compliance, Dryad product developments and events, and more. To get the latest news and developments from Dryad delivered straight to your inbox, please sign up to ... [Continue reading →](#)

Posted in [Quarterly update](#), [Uncategorized](#) | Tagged [AAAS](#), [Data Management](#), [data sharing](#), [data sharing policy](#), [Dryad](#), [open data](#), [Open Science](#), [Peer Review](#)

Quarterly roundup: Your February 2023 news from Dryad

Posted on 2023/02/06 by Sarah Lippincott

Welcome to the Dryad newsletter, keeping you up to date with data policies and compliance, Dryad product developments and events, and more. To get the latest news and developments from Dryad delivered straight to your inbox, please sign up to our ... [Continue reading →](#)

Posted in [Quarterly update](#) | Tagged [Dryad](#), [eJournalPress](#), [Library](#), [Newsletter](#), [Open House](#), [Open Science](#), [OSTP](#), [RDM](#), [SCOSS](#), [White House](#)

Quarterly roundup: Your November 2022 news from Dryad

Posted on 2022/11/09 by Sarah Lippincott

Welcome to the Dryad newsletter, keeping you up to date with data policies and compliance, Dryad product developments and events, and more. To get latest news and developments from

EMAIL SUBSCRIPTION

Email Address

Sign me up!

Join 12,579 other subscribers

Dryad on Twitter

ARCHIVES

- Data availability
- Data curation
- Data management plans
- Data reuse
- Discoverability
- Featured data
- Funding
- Governance
- Infrastructure
- Institutional
- Journals
- Meetings
- Membership
- New features
- Open Infrastructure
- Opportunities
- Partnerships
- Policy
- Product
- Publishers
- Quarterly update





CONNECTING RESEARCH,
IDENTIFYING KNOWLEDGE

Guide for funding organizations To support FAIR Workflows and enable output tracking

Xiaoli Chen

2023 September 7

Measuring the Impact of Data Sharing - HRA Webinar



[@datacite](https://twitter.com/datacite)



[@datacite@openbiblio.social](https://openbiblio.social/@datacite)



About DataCite and community



Global non-profit membership organization working with 2700+ repositories in the world to provide DOIs for research outputs and resources.



2800+

Repositories



280+

Members



51

Countries



52m+

DOIs



1200+

Organizations

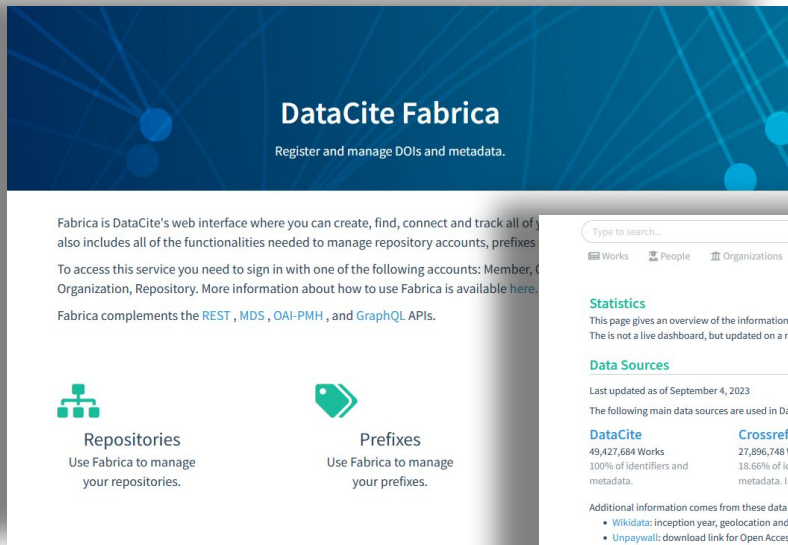
DataCite Services

Create, manage, and discover DOIs



DataCite makes research more effective by connecting research outputs and resources—from data and preprints to images and samples.

We support the **creation and management of DOIs** and **metadata records**, enhance research workflows with **service integration**, and **enable the discovery** and reuse of research outputs and resources.




DataCite Fabrica

Register and manage DOIs and metadata.


Fabrica is DataCite's web interface where you can create, find, connect and track all of your DOIs. It also includes all of the functionalities needed to manage repository accounts, prefixes and metadata.

To access this service you need to sign in with one of the following accounts: Member, Organization, Repository. More information about how to use Fabrica is available here.

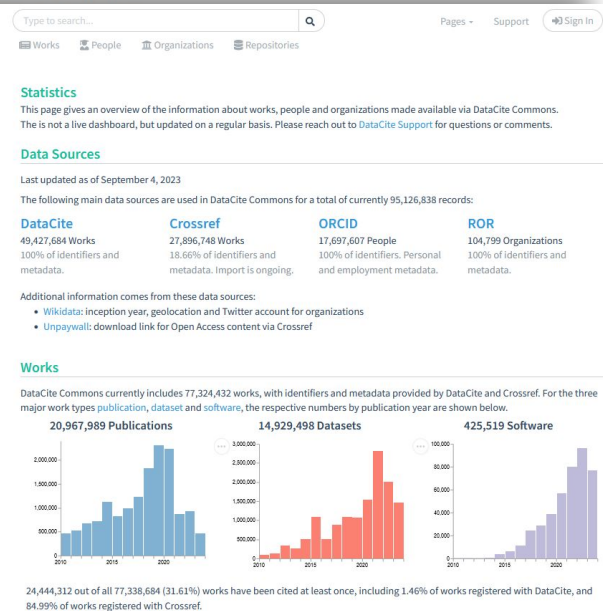
Fabrica complements the [REST](#), [MDS](#), [OAI-PMH](#), and [GraphQL](#) APIs.



Repositories
Use Fabrica to manage your repositories.



Prefixes
Use Fabrica to manage your prefixes.



Type to search...

Pages - Support Sign In

Works People Organizations Repositories

Statistics

This page gives an overview of the information about works, people and organizations made available via DataCite Commons. The is not a live dashboard, but updated on a regular basis. Please reach out to [DataCite Support](#) for questions or comments.

Data Sources

Last updated as of September 4, 2023

The following main data sources are used in DataCite Commons for a total of currently 95,126,838 records:

DataCite	Crossref	ORCID	ROR
49,427,684 Works 100% of identifiers and metadata.	27,896,748 Works 18.66% of identifiers and metadata. Import is ongoing.	17,697,607 People 100% of identifiers. Personal and employment metadata.	104,799 Organizations 100% of identifiers and metadata.

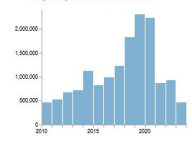
Additional information comes from these data sources:

- Wikidata: inception year, geolocation and Twitter account for organizations
- Unpaywall: download link for Open Access content via Crossref

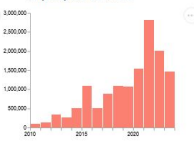
Works

DataCite Commons currently includes 77,324,432 works, with identifiers and metadata provided by DataCite and Crossref. For the three major work types *publication*, *dataset* and *software*, the respective numbers by publication year are shown below.

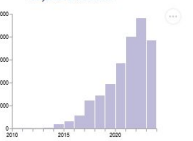
20,967,989 Publications



14,929,498 Datasets



425,519 Software



24,444,312 out of all 77,338,684 (31.61%) works have been cited at least once, including 1.46% of works registered with DataCite, and 84.99% of works registered with Crossref.

Resource types supported By the DataCite metadata schema

The DataCite Metadata Schema accommodates many different types of scholarly resources.

For any of these resource types, a funding reference can be added to its metadata record, when registering a DOI.

Audiovisual
Book
BookChapter
Collection
ComputationalNotebook
ConferencePaper
ConferenceProceeding
DataPaper
Dataset
Dissertation
Event
Image
InteractiveResource
Journal

JournalArticle
Model
OutputManagementPlan
PeerReview
PhysicalObject
Preprint
Report
Service
Software
Sound
Standard
Text
Workflow
Other

Funding reference in DataCite Schema

1. Identifier
2. Creator
3. Title
4. Publisher
5. PublicationYear
6. Subject
7. Contributor
8. Date
9. Language
10. ResourceType
11. AlternateIdentifier
12. RelatedIdentifier
13. Size
14. Format
15. Version
16. Rights
17. Description
18. GeoLocation
19. FundingReference
20. RelatedItem

19. FundingReference

- Definition: Information about financial support (funding) for the resource being registered.
- It is a best practice to supply funding information when financial support has been received.

Sub-properties:

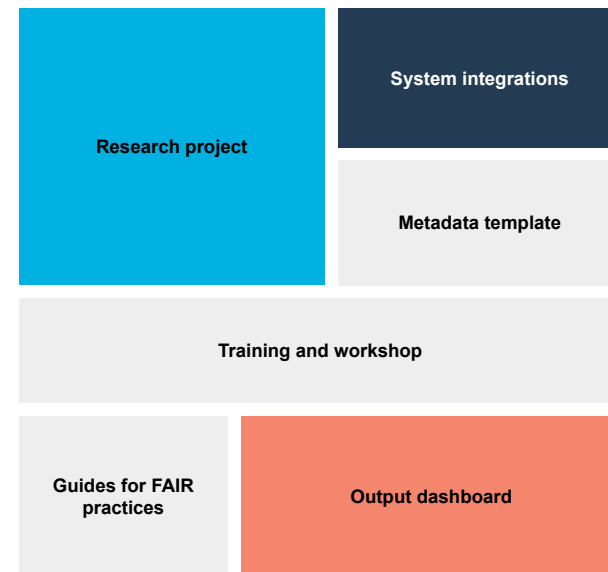
- 19.1 funderName
- 19.2 funderIdentifier
 - 19.2.a funderIdentifierType
 - 19.2.b schemeURI
- 19.3 awardNumber
 - 19.3.a awardURI
- 19.4 awardTitle

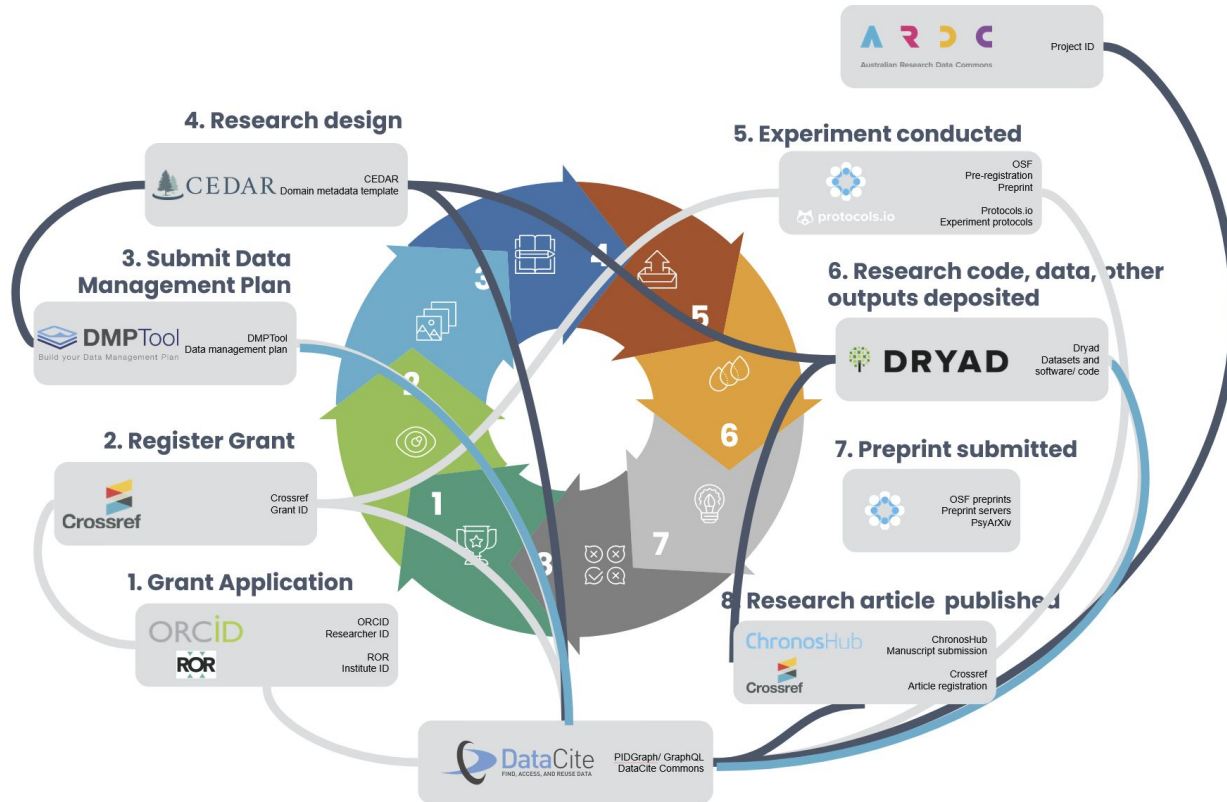
```
<fundingReferences>
  <fundingReference>
    <funderName>European Commission</funderName>
    <funderIdentifier funderIdentifierType="Crossref Funder
ID">https://doi.org/10.13039/501100000780</funderIdentifier
>
    <awardNumber
awardURI="https://cordis.europa.eu/project/rcn/100180_en.ht
ml">282625</awardNumber>
    <awardTitle>MOTivational strength of ecosystem services
and alternative ways to express the value of
BIODiversity</awardTitle>
  </fundingReference>

  <fundingReference>
    <funderName>European Commission</funderName>
    <funderIdentifier funderIdentifierType="Crossref Funder
ID">https://doi.org/10.13039/501100000780</funderIdentifier
>
    <awardNumber
awardURI="https://cordis.europa.eu/project/rcn/100603_en.ht
ml">284382</awardNumber>
    <awardTitle>Institutionalizing global genetic-resource
commons. Global Strategies for accessing and using
essential public knowledge assets in the life
sciences</awardTitle>
  </fundingReference>
</fundingReferences>
```

Implementing FAIR Workflows

A Proof of Concept Study in the Field of Consciousness





Guide for Funders to Support FAIR Workflows and Enable Output Tracking



Crossref

URL form

Prefix

[https://doi.org/10.5281/](https://doi.org/10.5281/zenodo.8289142)

[zenodo.8289142](https://doi.org/10.5281/zenodo.8289142)

Suffix



Project Title **Implementing FAIR Workflows: a proof of concept study in the field of consciousness**

TWCF Grant No **TWCF0568** <https://doi.org/10.54224/20568>

D3.2 Guide for funders to support FAIR workflows & enable research tracking

Date **August 31, 2023**

Authors **Xiaoli Chen, DataCite
Helena Cousijn, DataCite
Ginny Hendricks, Crossref
Shawna Sadler, ORCID
Kelly Stathis, DataCite**

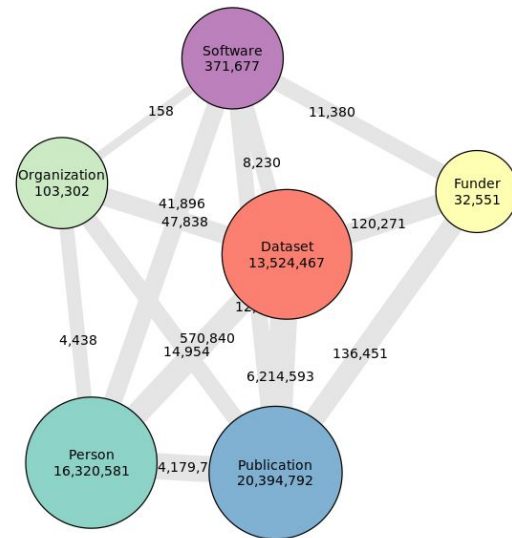
Status **Public**



This project was made possible through the support of a grant from Templeton World Charity Foundation, Inc. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of Templeton World Charity Foundation, Inc.

The recommendations in a nutshell

“Registering PIDs for outputs, and associating them on the metadata level with the PIDs of the grant, the funding organization and the fundees is the basic logic behind a robust and automated output tracking mechanism.”



**Know and maintain
funder IDs**

**Create and manage
grant IDs**

**Provide guidance for
researchers and other
stakeholders**



Know and maintain funder identifiers

Knowing the funder ID

- Crossref Open Funder Registry (OFR) ID
- Research Organization Registry (ROR) ID

Maintaining the funder ID

- Keep the metadata record up-to-date
 - Capture merging, combining, re-organising departments, or rebranding of the organization
- Update policies and requirement documents to communicate the existence and use case of funder ID to relevant parties

	Funder ID	Metadata via API
	https://doi.org/10.13039/501100011730	https://api.crossref.org/funders/501100011730
	https://ror.org/00x0z1472	https://api.ror.org/organization/00x0z1472

Create and manage grant identifiers

Grant IDs are fully within the funder's control and responsibility.

- Create metadata record for grants, include information about:
 - Grant award
 - Funded project(s)
 - Funded investigator(s)
 - Role and affiliation
 - Publicly accessible landing page
- Integrate Grant ID management workflow based on output tracking needs
 - Create and maintain multilateral links

Funding types supported by Crossref grant ID schema

1. **award**: a prize, award, or other type of general funding
2. **contract**: agreement involving payment
3. **crowdfunding**: funding raised via multiple sources, typically small amounts raised online
4. **endowment**: gift of money that will provide an income
5. **equipment**: use of or gift of equipment
6. **facilities**: use of location, equipment, or other resources
7. **fellowship**: grant given for research or study
8. **grant**: a monetary award
9. **loan**: money or other resource given in anticipation of repayment
10. **other**: award of undefined type
11. **prize**: an award given for achievement
12. **salary-award**: an award given as salary, includes intramural research funding
13. **secondment**: detachment of a person or resource for temporary assignment elsewhere
14. **seed-funding**: an investor invests capital in exchange for equity
15. **training-grant**: grant given for training

Provide guidance for researchers

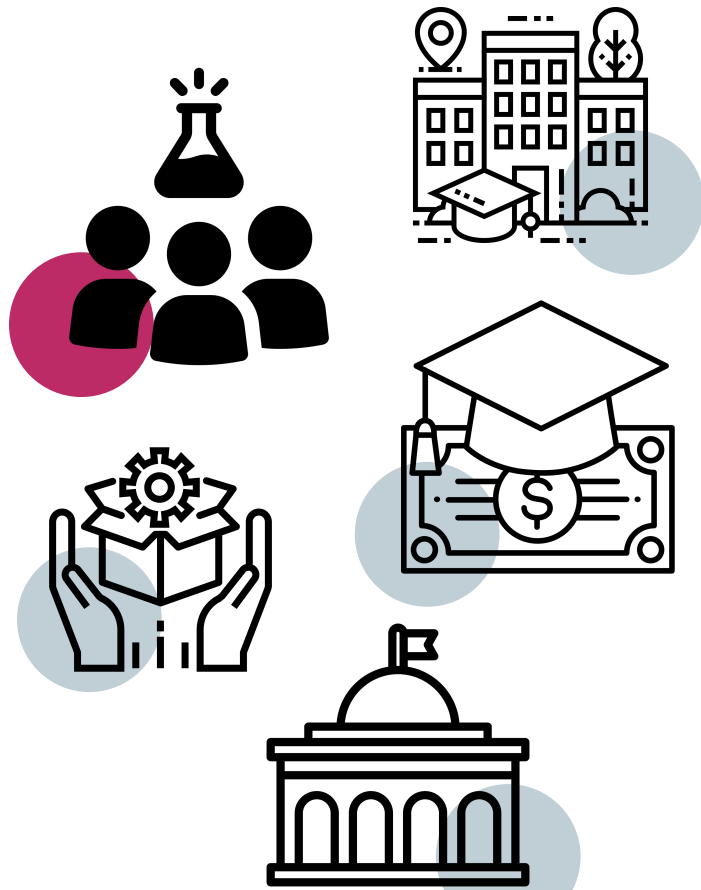
Identifier for researchers



- Metadata of ORCID include information about researcher's employment, funding, and outputs
- Require grant applicants to create and use ORCID iDs

Incentivizing researchers

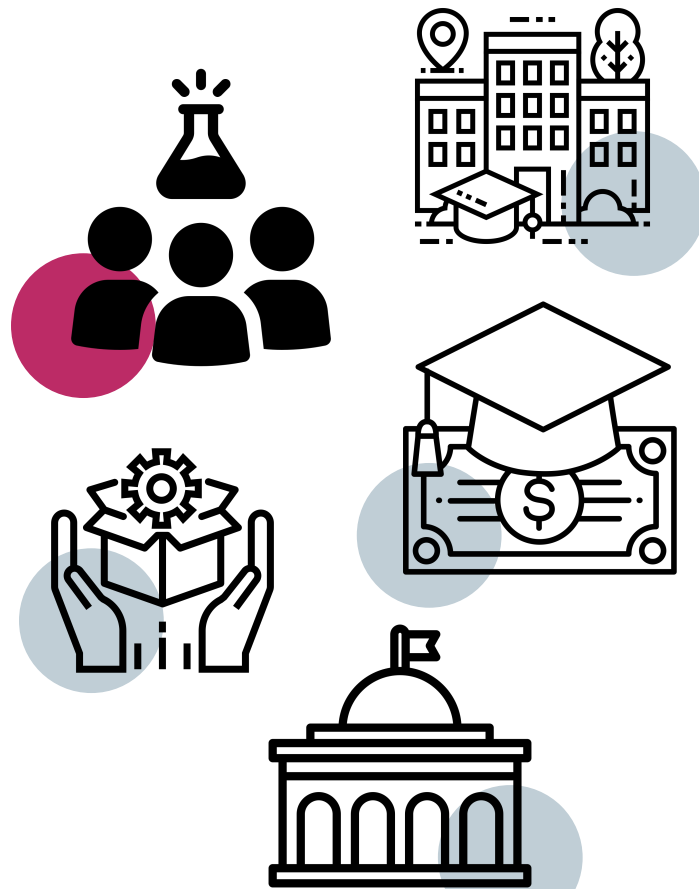
- Include the Open and FAIR sharing track record of applicants as part of evaluation criteria for grant project proposals
- Provide guidance for open sharing and request data management plan for grant applications
- Setting aside specific budget and allowing time in the project schedule to accommodate the incorporation of robust FAIR workflows
- Promoting reusable outputs to increase visibility of researchers and research teams that practice FAIR



Provide guidance for researchers

Data Management Planning

- Request or mandate data management plan as part of the grant application portfolio
- Incorporate data management planning policy into grant agreement
- If mandating a DMP, consider providing a DMP template based on output tracking needs.
- Include compliance monitoring mechanism in the DMP policy.
- Instruct researchers to include ORCID, ROR ID, grant ID, and funder ID in the DMP metadata.
- Recommend researcher to use DMP platform that supports machine readable DMP format and registers PIDs for DMPs.



Engaging with other stakeholders

Create the condition for the various PID and metadata workflows that can be adopted by publishers, institutes, policy makers, and other funding organizations.

- Encourage technology providers to integrate open scholarly infrastructure
- Align with other funding organizations on best practices around PIDs and metadata
- Work with publishers to enhance their PID workflows to include grant and funder metadata
- Engage with government agencies to steer policy making



Please share your thoughts!

[Open for comments \(September 1 - October 15\)](#)



CONNECTING RESEARCH,
IDENTIFYING KNOWLEDGE



info@datacite.org



pidforum.org



datacite.org
blog.datacite.org



support@datacite.org
support@datacite.org



[@datacite](https://twitter.com/datacite)



[DataCite](https://www.youtube.com/DataCite)



[@datacite](https://www.linkedin.com/company/datacite)



CONNECTING RESEARCH,
IDENTIFYING KNOWLEDGE

It's time to Make Data Count

Iratxe Puebla
Director, Make Data Count

7 September 2023
Measuring the Impact of Data Sharing - HRA Webinar



[@datacite](https://twitter.com/datacite)
[@makedatacount](https://twitter.com/makedatacount)




[@datacite@openbiblio.social](https://openbiblio.org/@datacite@openbiblio.social)
[@makedatacount@openbiblio.social](https://openbiblio.org/@makedatacount@openbiblio.social)



Data sharing is valuable, but do we understand the value of data sharing?



 EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

August 25, 2022

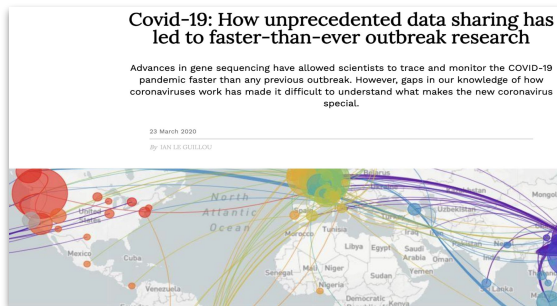
MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Dr. Alondra Nelson *Alondra Nelson*
Deputy Assistant to the President and Deputy Director for Science and Society
Performing the Duties of Director
Office of Science and Technology Policy (OSTP)

SUBJECT: Ensuring Free, Immediate, and Equitable Access to Federally Funded Research

This memorandum provides policy guidance to federal agencies with research and development expenditures on updating their public access policies. In accordance with this memorandum, OSTP recommends that federal agencies, to the extent consistent with applicable law:

1. Update their public access policies as soon as possible, and no later than December 31st, 2025, to make publications and their supporting data resulting from federally funded research publicly accessible without an embargo on their free and public release;
2. Establish transparent procedures that ensure scientific and research integrity is maintained in public access policies; and,
3. Coordinate with OSTP to ensure equitable delivery of federally funded research results and data.



OPEN WHEAT BLAST

MAKING DATA INSTANTLY ACCESSIBLE

CONSEQUENCES OF INACTION OUR MISSION DATA DOWNLOAD BIOCONTROL BIOMATERIALS
DATA ANALYSIS HOW YOU CAN HELP RESOURCES WHO WE ARE CONTACT US

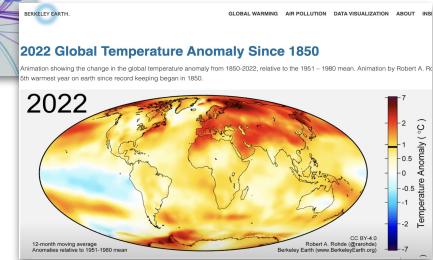
LET'S MAKE A DIFFERENCE

Type here to search...


VISIT OUR SISTER WEBSITE
openwheatblast.net

Wheat blast is a fearsome fungal disease of wheat. It is caused by a fungus known as *Magnaporthe oryzae* (syn. *Puccinia oryzae*). There is a risk that wheat blast could spread beyond South America and threaten food security in wheat growing areas in Asia and Africa.

OPEN WHEAT BLAST



Data Sharing: A Roadmap



Is your organization considering adopting a data sharing policy (DSP), but you're not sure where to begin? HRA members have built this guide to provide an overview of the process and highlight available resources.

Note: The steps outlined below are not necessarily linear - you may find that a different sequence is more conducive to your organization's goals, requirements, or timeline.

Explore and research

Determine whether it's the right time. Do you have the resources and support to undertake a large organizational change?

- Do the grants your organization fund generate big data (e.g. genomic databases, clinical trial databases, imaging databases, computational models)?
- Do you have enough administrative and programmatic support to adjust your grant practices to incorporate a DSP?

Consider taking incremental steps before/while your organization develops its own policy:

- Ask awardees to submit a data sharing plan absent of incentive and without encouragement in a particular direction.
- Add encouragement by stating that the organization wants data sharing, and that applications including data sharing plans that promote broad, rapid sharing will be scored favorably.
- Ultimately, your organization may require, score, and benchmark applications based on strong data sharing plans.

Understanding the impact of open data requires transparent and responsible data metrics

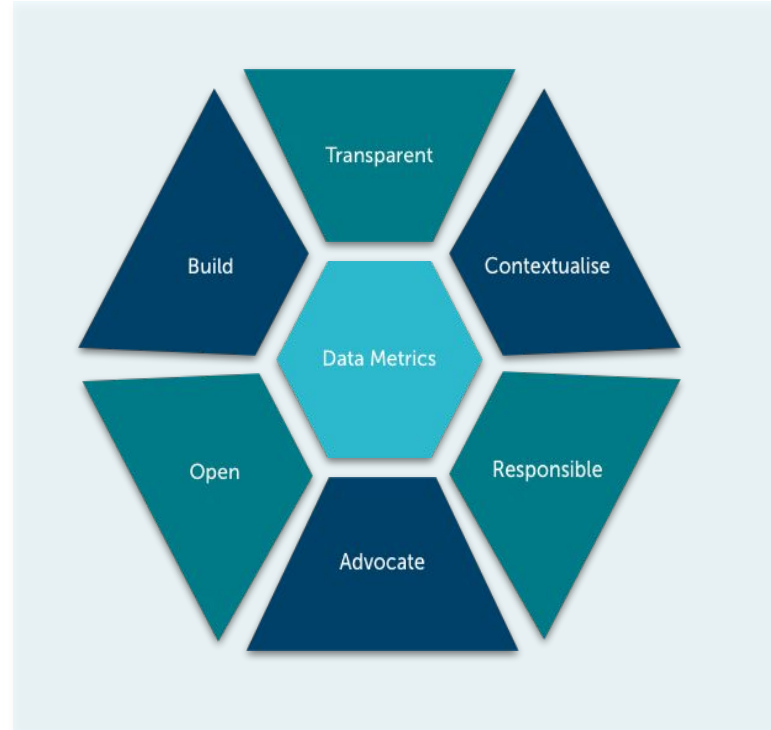
Make Data Count

Make Data Count is an initiative that **promotes open data metrics** to enable **evaluation and reward of research data** reuse and impact.

Community effort to ensure that data are used and cited in open, transparent, and responsible ways.

- **Build** open infrastructure and community-based standards.
- **Advocate** through outreach and adoption campaigns.
- **Contextualize** with evidence-based bibliometric studies.

makedatacount.org/about-us/



Data Citations

Data citations are references to data, in the same way researchers routinely provide a bibliographic reference to other scholarly resources.

Data citations do not provide the full picture of data usage, but they are useful as they clearly point to data being used or reused in research.

SampleID	Individual_ID	UniProt	GeneID	Assay	NPX	Panel	Index
CV21_1	CV21	P42785	PRCP	PRCP	2.20938	CM	1
C161_2	C161	P42785	PRCP	PRCP	1.58386	CM	2
C173_3	C173	P42785	PRCP	PRCP	0.85673	CM	3
CV11_4	CV11	P42785	PRCP	PRCP	1.63151	CM	4
CV39_5	CV39	P42785	PRCP	PRCP	0.64489	CM	5
C153_6	C153	P42785	PRCP	PRCP	0.65126	CM	6
C33_7	C33	P42785	PRCP	PRCP	0.69374	CM	7
CV39_8	CV39	P42785	PRCP	PRCP	1.09712	CM	8
CV6_9	CV6	P42785	PRCP	PRCP	1.8895	CM	9
CV38_10	CV38	P42785	PRCP	PRCP	1.72346	CM	10
CV4_11	CV4	P42785	PRCP	PRCP	0.69872	CM	11
C195_12	C195	P42785	PRCP	PRCP	1.07357	CM	12
CV1_13	CV1	P42785	PRCP	PRCP	0.61055	CM	13
C237_14	C237	P42785	PRCP	PRCP	1.19429	CM	14
CV2_15	CV2	P42785	PRCP	PRCP	0.32423	CM	15
C189_16	C189	P42785	PRCP	PRCP	0.94542	CM	16
CV42_17	CV42	P42785	PRCP	PRCP	1.97023	CM	17
CV79_18	CV79	P42785	PRCP	PRCP	1.38031	CM	18
CV3_19	CV3	P42785	PRCP	PRCP	0.28226	CM	19
CV36_20	CV36	P42785	PRCP	PRCP	1.41429	CM	20
CV55_21	CV55	P42785	PRCP	PRCP	1.3812	CM	21
CV46_22	CV46	P42785	PRCP	PRCP	1.18128	CM	22
CV30_23	CV30	P42785	PRCP	PRCP	0.72615	CM	23

Gisby J, Clarke CL, Medjeral-Thomas N, Malik TH, Papadaki A, Mortimer PM, Buang NB, Lewis S, Pereira M, Toulza F, Fagnano E, Mawhin M, Dutton EE, Tapeng L, Kirk P, Behmoaras J, Sandhu E, McAdoo SP, Predecki MF, Pickering MC, Botto M, Willicombe W, Thomas DC, Peters JE (2020) **Dryad Digital Repository** Longitudinal proteomic profiling of high-risk patients with COVID-19 reveals markers of severity and predictors of fatal disease.

<https://doi.org/10.5061/dryad.6t1g1jwuj>

eLife RESEARCH ARTICLE

Longitudinal proteomic profiling of dialysis patients with COVID-19 reveals markers of severity and predictors of death

Jack Gisby^{1†}, Candice L Clarke^{1,2†}, Nicholas Medjeral-Thomas^{1,2†}, Talat H Malik¹, Artemis Papadaki¹, Paige M Mortimer¹, Norzawan B Buang¹, Shanice Lewis¹, Marie Pereira¹, Frederic Toulza¹, Ester Fagnano¹, Marie-Anne Mawhin¹, Emma E Dutton¹, Lunnathaya Tapeng¹, Arianne C Richard^{3,4}, Paul DW Kirk^{5,6,7}, Jacques Behmoaras¹, Eleanor Sandhu¹, Stephen P McAdoo^{1,8}, Maria F Predecki^{1,9}, Matthew C Pickering¹, Marina Botto¹, Michelle Willicombe^{1,10}, David C Thomas^{1,2,11}, James E Peters^{1,2,11*}

¹Centre for Inflammatory Disease, Department of Immunology and Inflammation, Imperial College London, London, United Kingdom; ²Renal and Transplant Centre, Hammersmith Hospital, Imperial College Healthcare NHS Trust, London, United Kingdom; ³Cambridge Institute for Medical Research, University of Cambridge, Cambridge, United Kingdom; ⁴CRUK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom; ⁵MRC Biostatistics Unit, Forvie Way, University of Cambridge, Cambridge, United Kingdom; ⁶Cambridge Institute of Therapeutic Immunology & Infectious Disease, University of Cambridge, Cambridge, United Kingdom; ⁷Health Data Research UK, London, United Kingdom

*For correspondence: j.peters@imperial.ac.uk

[†]These authors contributed equally to this work

[‡]These authors also contributed equally to this work

Competing Interest: See [page 24](#)

Funding: See [page 25](#)

Received: 12 November 2020
Accepted: 10 March 2021
Published: 11 March 2021

Reviewing editor: Evangelos J Giannellis-Bourbouts, National

Abstract: End-stage kidney disease (ESKD) patients are at high risk of severe COVID-19. We measured 436 circulating proteins in serial blood samples from hospitalised and non-hospitalised ESKD patients with COVID-19 (n = 256 samples from 55 patients). Comparison to 51 non-infected patients revealed 221 differentially expressed proteins, with consistent results in a separate cohort of 46 COVID-19 patients. Two hundred and three proteins were associated with clinical severity, including IL6, markers of monocyte recruitment (e.g. CCL2, CCL7), neutrophil activation (e.g. proteinase-3, and epithelial injury (e.g. KRT19). Machine-learning identified predictors of severity including IL18BP, CTSD, GDF15, and KRT19. Survival analysis with joint models revealed 69 predictors of death. Longitudinal modelling with linear mixed models uncovered 32 proteins displaying different temporal profiles in severe versus non-severe disease, including integrins and adhesion molecules. These data implicate epithelial damage, innate immune activation, and leucocyte-endothelial interactions in the pathology of severe COVID-19 and provide a resource for identifying drug targets.

Data Citations

Data Citations enable

Credit for the researchers who produce the data

Transparency & reproducibility for those using the data

Evaluation of open data

<https://commons.datacite.org/doi.org/10.5061/dryad.cz8w9qj17>



DataCite Commons

[Add to ORCID Record](#)

[Download Metadata](#)

Cite as

Lyon, A., Rugema, N., Garland-Kuntz, E., Sieng, M., Muralidharan, K., Van Camp, M. M., O'Neill, H., Mbongo, W., Selvia, A. F., Marti, A. T., Everly, A., McKenzie, E., & Lyon, A. M. (2020). *Structure of phospholipase Cε reveals an integrated RA1 domain and previously unidentified regulatory elements (Version 6)* [Data set]. Dryad. <https://doi.org/10.5061/DRYAD.CZ8W9GJ17>

APA

Share

Email

Twitter

Facebook

Filter Works

Type to search... x Q

Publication Year

<input type="checkbox"/> 2020	1
<input type="checkbox"/> 2018	2
<input type="checkbox"/> 2007	1

Publication Year

Work Types

Text 100%

Licenses

Missing 75%

Structure of phospholipase Cε reveals an integrated RA1 domain and previously unidentified regulatory elements

<https://doi.org/10.5061/dryad.cz8w9qj17>

4 Citations 73 Views 2 Downloads

Description Creators Funders Registration

Phospholipase Cepsilon (PLCepsilon) generates lipid-derived second messengers in the cardiovascular system at the plasma and perinuclear membranes. It is activated in response to a wide variety of signals, such as those conveyed by Rap1A and Ras, through a mechanism that involves its C-terminal Ras association (RA) domains (RA1 and RA2). However, the complexity and size of PLCepsilon has hindered its structural and functional analysis. In this manuscript, we report the 2.7 Å crystal structure of fragment of PLCepsilon that retains catalytic activity. The structure includes the RA1 domain, which forms an integral part of the conserved core. In addition, a highly conserved amphipathic helix in the autoinhibitory X-Y linker is shown to modulate activity in vitro and in cells. The studies provide a structural framework for the core of this critical cardiovascular enzyme that will allow for a better understanding of its regulation and roles in disease.

Version 6 of Dataset published 2020 in DRYAD

Dataset English

<https://doi.org/10.5061/dryad.cz8w9qj17>

73 Views 2 Downloads

73 views reported since publication in 2020.

Data Citations: The challenges

- Researchers do not always cite the data they use.
- Data citations can appear at different locations in an article: Methods section, Data Availability statement, References, footnotes.
- Citations and/or associated metadata may be lost when the publisher deposits in Crossref.
- Many datasets in the life sciences and biomedical fields use accession numbers instead of DOIs.

We know there are many more instances of data usage than we are currently capturing

Data Citations: How funders can help

Funders can drive awareness about best practices in data usage among grantees and incentivize adoption by incorporating data sharing and reuse in evaluation processes



Encourage researchers to deposit datasets at repositories that assign DOIs



Encourage researchers to cite datasets they have used - their own and others' - in their research outputs

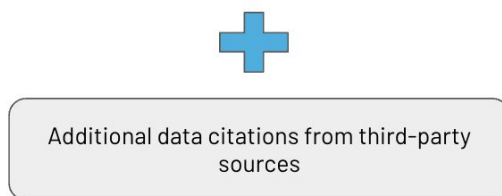
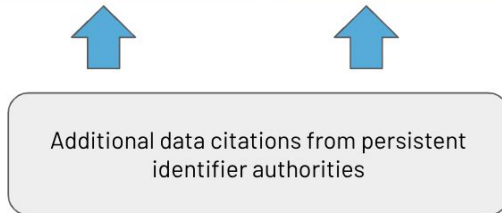


In grant applications and reports, ask researchers to report the datasets they shared and their reuse & consider this as part of evaluation processes

In addition to promoting best practices, we also need infrastructure and services that automate the data citation process and make metadata inclusion easy for researchers.

Global Open Data Citation Corpus

Goal: Develop a comprehensive corpus that incorporates data citations from different sources into a centralized, publicly accessible community resource



Incorporate data citations from diverse sources:

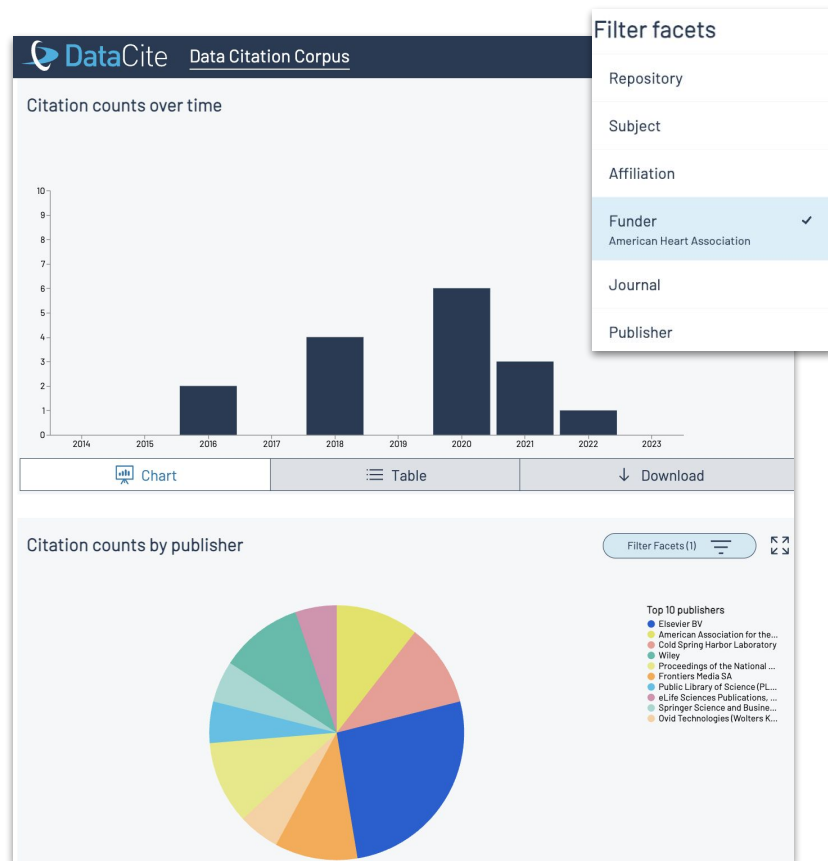
- Persistent Identifier (PID) authorities (e.g. Crossref, DataCite) that collect citations as part of their metadata deposit workflow.
- Additional sources that aggregate or discover citations through various techniques, such as machine learning and curation of full-text in articles.

Data Citation Corpus: Prototype

The prototype will make the initial seed data available and provide a basic user interface to visualize the data with different filters.

Our longer term goal is to enable stakeholders, including funders, to use the corpus as part of their processes to:

- Evaluate the outputs and reach of researchers' work, e.g. in grant applications or reports.
- Report on the reach of research outputs from funded projects.



Thank you!

Make Data Count Summit

September 12-13, Washington DC

Bringing together researchers, funders, government data administrators, publishers and infrastructure providers to discuss data impact, data usage, and data metrics.

summit.makedatacount.org

We will seek community input as we work on the development of the data citation corpus. Interested in learning more? Do get in touch:

iratxe.puebla@datacite.org
info@datacite.org

Data citation corpus:
makedatacount.org/data-citation

Data sharing workflow

Funding

- Preliminary data
- Data sharing plan

Research

- Dataset creation & collection
- Data storage

Analysis & Writing

- Preparing to share findings
- Data curation

Publishing/ Sharing

- Dataset in a FAIR repository
- Data citation

Post-Publication

- Data citation
- Re/use metrics



Funders,
Grantees

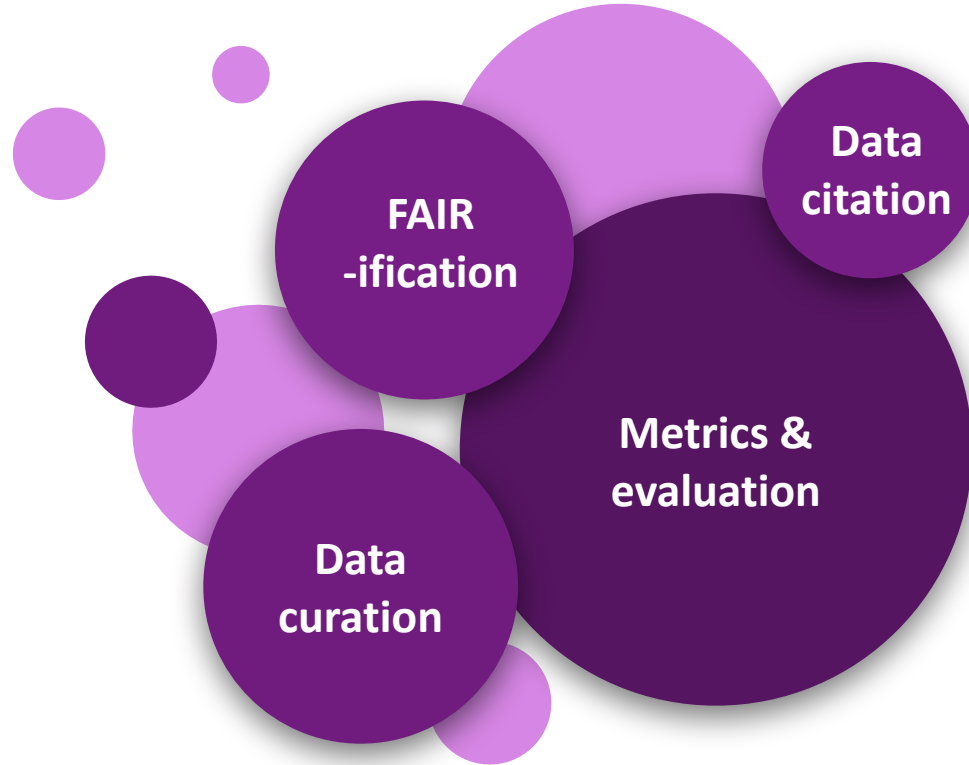
Researchers,
Institutions

Co-/Authors,
Data helpers

Publishers,
Infrastructures

Many
stakeholders

For discussion: What is the right toolchain?



Actions and Resources

[Fall 2023 Data Sharing BOF](#) Sign up sheet

[Updated Policy Worksheet](#) - save as your own working document

Email me with questions: kristen@strategiesos.org

Thank you!