

# Accelerating Reuse by Implementing FAIR Workflows

**Xiaoli Chen**

14 June, 2022

7th International Digital Curation Conference

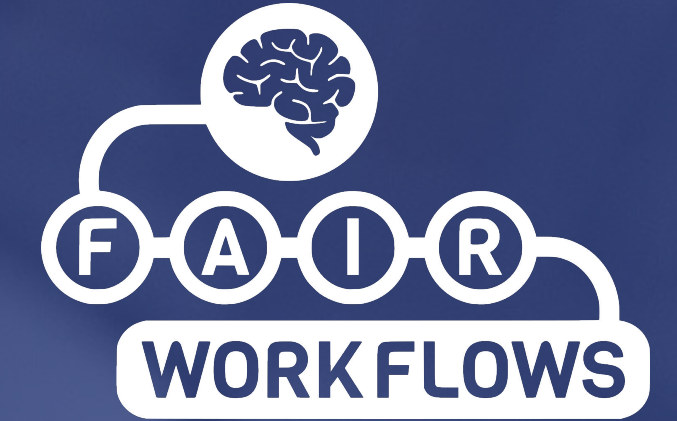
[xiaoli.chen@datacite.org](mailto:xiaoli.chen@datacite.org)

<https://twitter.com/DataCite>



# Meeting Agenda

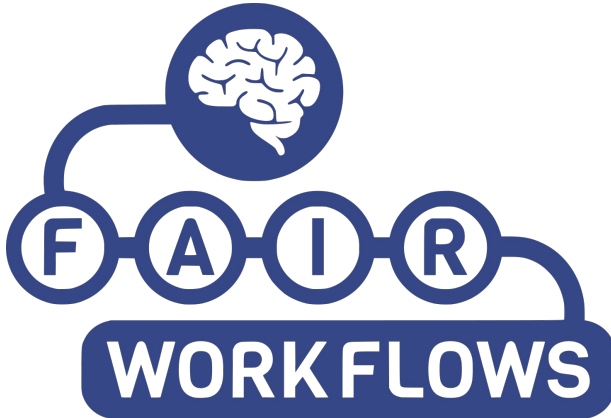
- A closer look at reuse
- The FAIR Workflows Project
- Progress and next steps



# A Closer Look at Reuse



# A Closer Look at Reuse



## Scientists'

Youngseek Kim  
School of Information  
E-mail: youngseek.ki

Ayoung Yoon  
School of Informatics  
Purdue University, 53

This study explores the reuse behaviors of scientists across different disciplines. This research uses a theoretical framework combining the theory of planned behavior and the theory of diffusion of innovation to test the proposed model. Study participants included 1,528 scientists from various disciplines. The analysis method was a structural equation model. The results show that several factors influence scientists' data reuse intentions, including perceived usefulness, perceived effort, and internal resources. Three main areas that influence data reuse intentions are identified: Research context, Data context and entity, and Behavior.

**Introduction**

Data sharing and reuse are essential for advancing science, scholarship, and societal well-being. Opportunities, validation of ex

Additional Supporting Information for this article

TABLE 4 Results of axial coding

Dimension	Main category	Category
Cognition	Make a decision	A1 attitudes
		A2 data reusability assessment
		A3 disciplinary climate
		A4 external assistance availability
		A5 intentions of data reuse
Data context and entity	Discover and acquire data	A6 data source
		A7 scaffold helping discover data
		A8 data discovery method
		A9 data access method
		A10 barrier preventing data discovery or access
	Understand and choose data	A11 understanding data properties
		A12 scaffold helping understand data
		A13 method to get scaffold helping data understanding
		A14 barriers preventing data understanding
		A15 processing data for research needs
		A16 barrier preventing data processing and reusing
		A17 an iterative process
Research context	Process and reuse data	
Behavior	The iterative property	

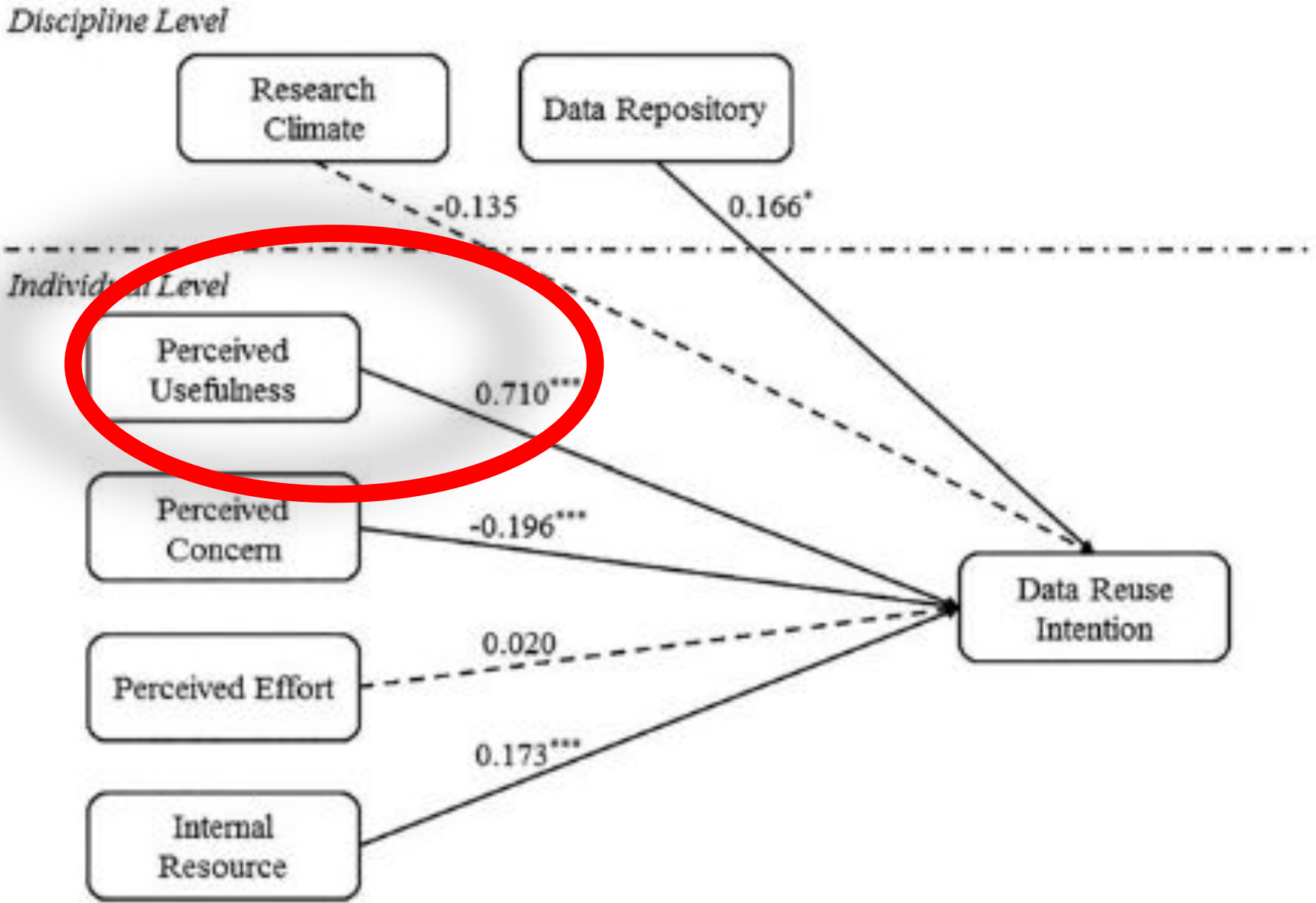


FIG. 2. Hypothesis testing results based on scientists' data reuse intention model.

Keywords: reuse, research data, open science, definition, characteristics

How to Cite: van de Sandt, S., Dallmeier-Tiessen, S., Lavasa, A. and Petras, V., 2019. The Definition of Reuse. *Data Science Journal*, 18(1), p.22. DOI: <http://doi.org/10.5334/dsj-2019-022>



# A Closer Look at Reuse



## **Reuse require deliberation, preparation, and facilitation**

- Researchers need to be informed of and committed to reuse, and produce reusable resources.
- It takes foresight, planning and continuous upkeep to ensure the reusability of research components and outputs.
- Infrastructure and services that guide and facilitate significantly ease the process.

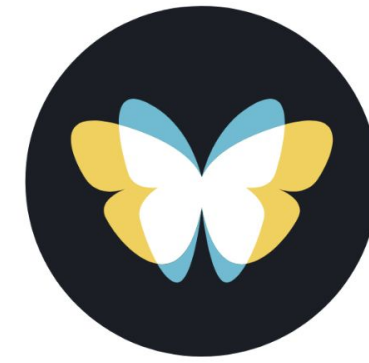
**Implementing FAIR Workflows**

**a proof of concept study in the  
field of consciousness**

# Project motivation

- Papers are not enough
- Aggravating reproducibility problems
- Research infrastructure and tools exist but adoption fragmented
- Researchers need concrete examples

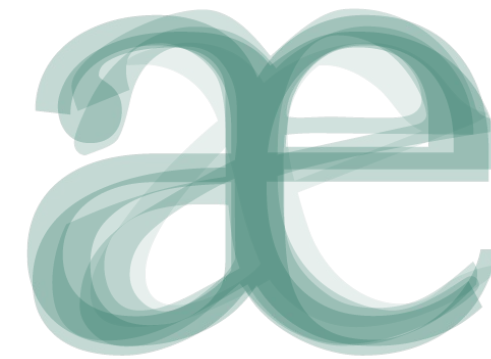




TEMPLETON WORLD  
CHARITY FOUNDATION



DataCite  
FIND, ACCESS, AND REUSE DATA



Max-Planck-Institut  
für empirische Ästhetik



University of California  
CDL  
California Digital Library



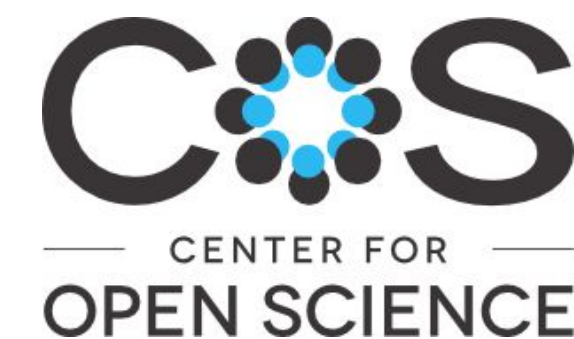
Australian Research Data Commons



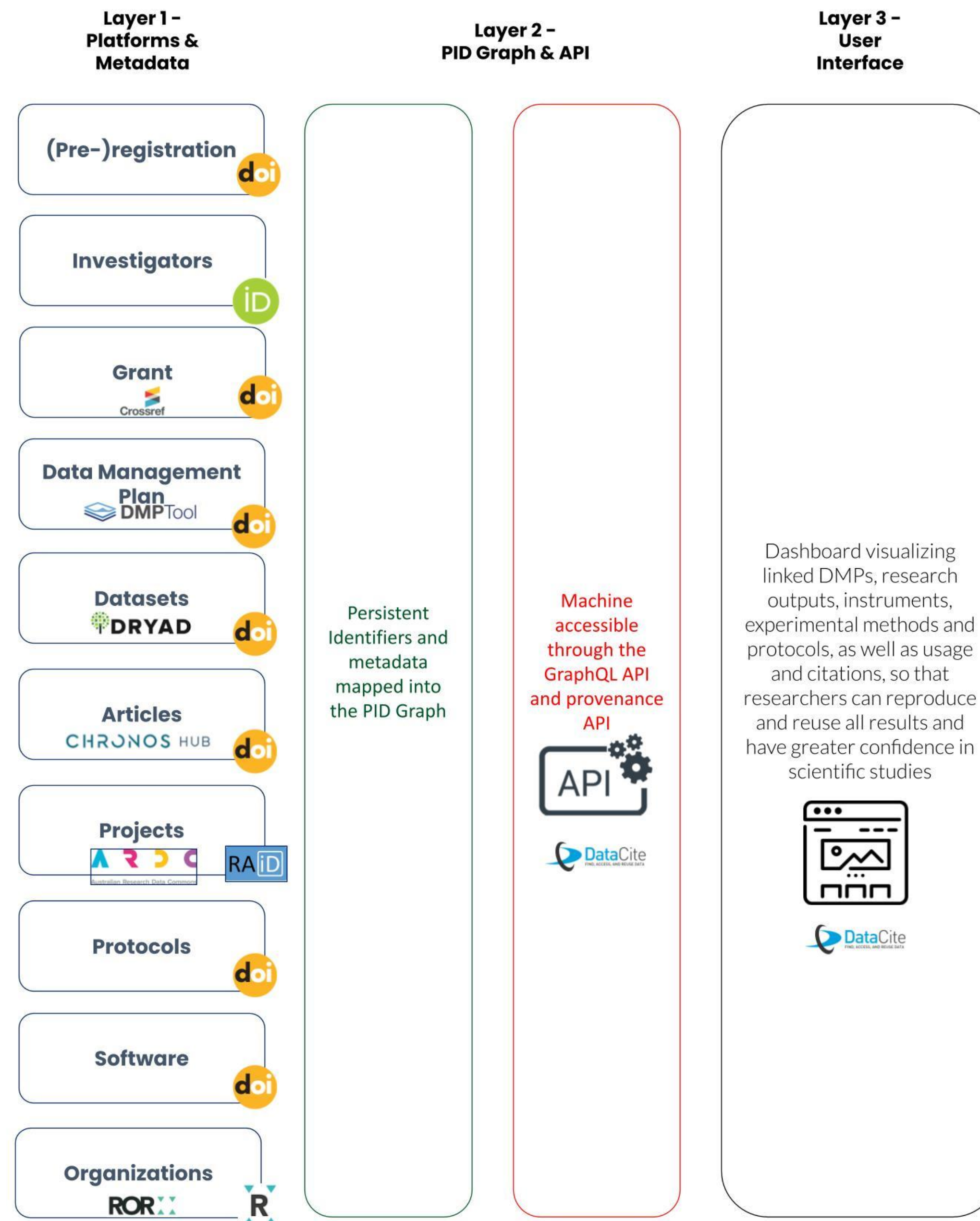
DRYAD

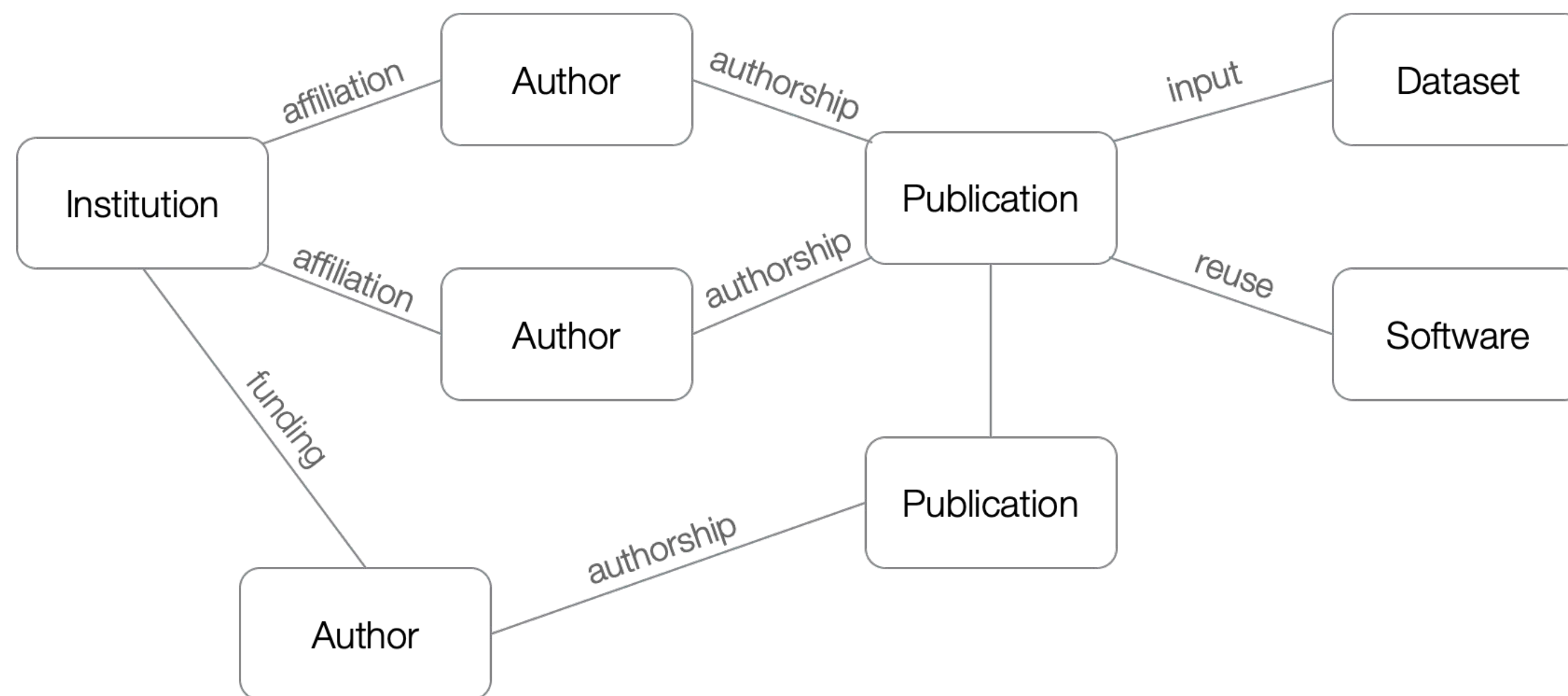
ChronosHub

Stanford  
University



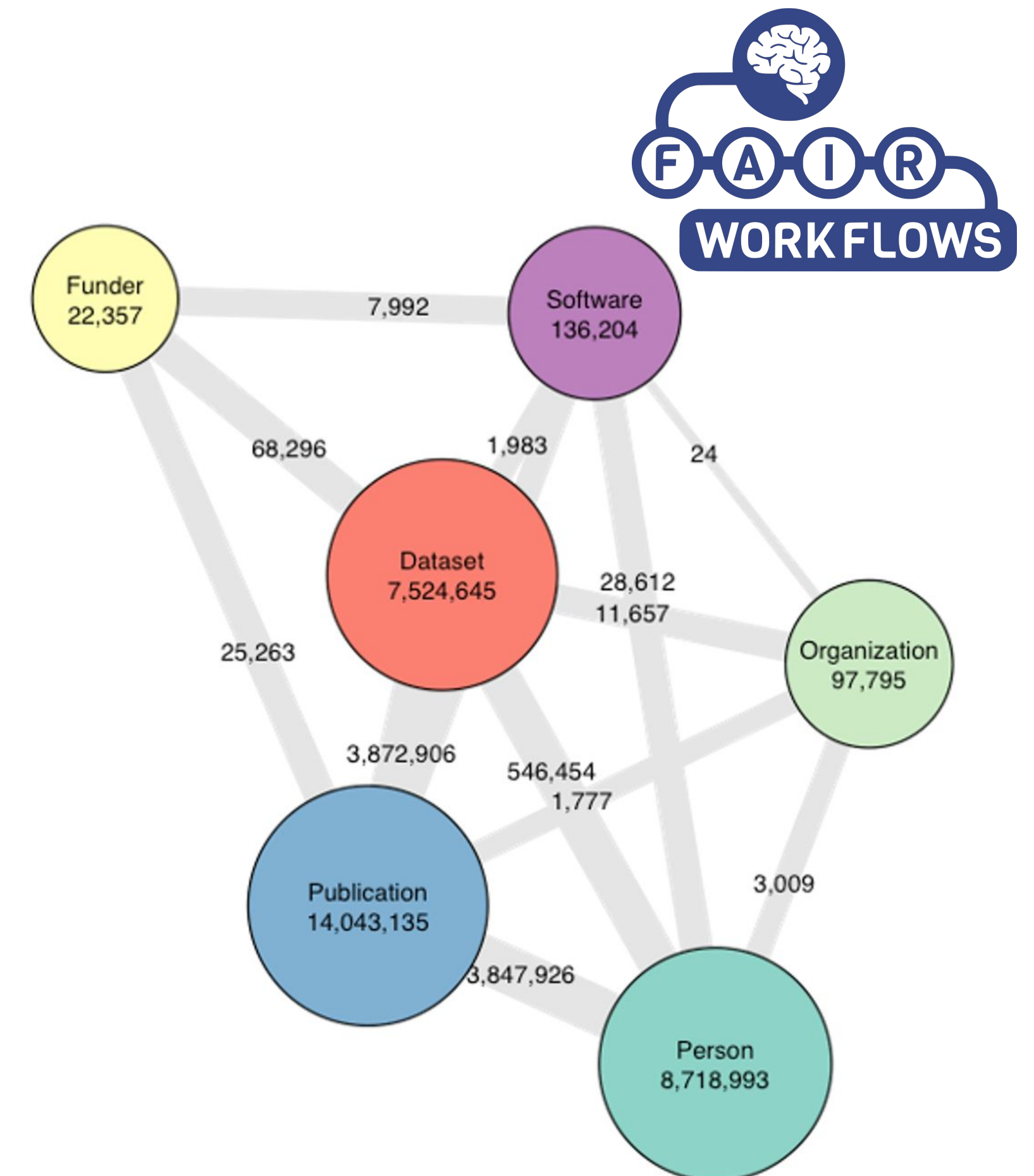






## Metadata properties

and their relations



## PID Graph

Numbers of nodes and connections  
(May 2020)

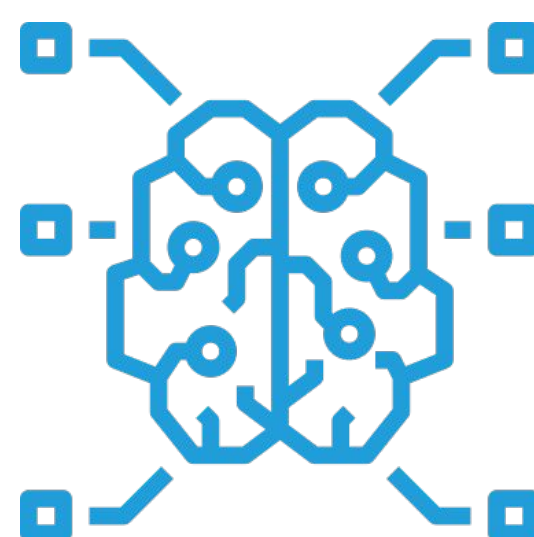


# Progress and next steps

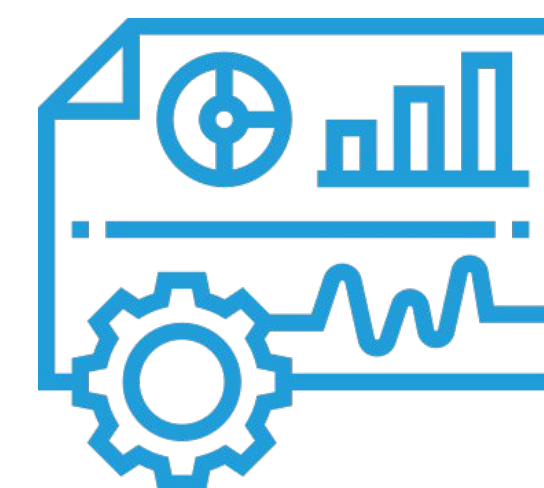
# Project approach



**Workflows  
development**



**Integration/  
application**



**PID graph &  
dashboard**



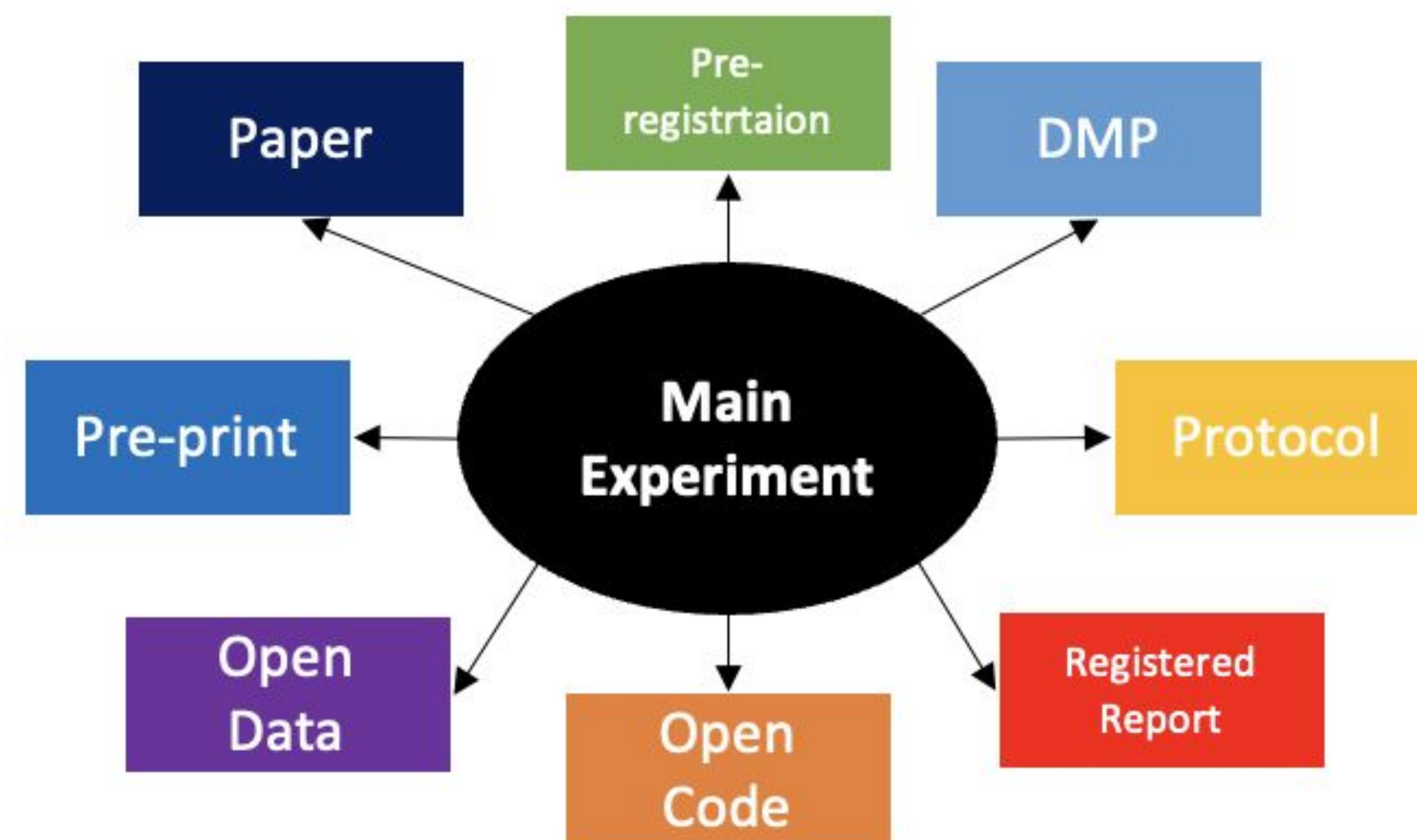
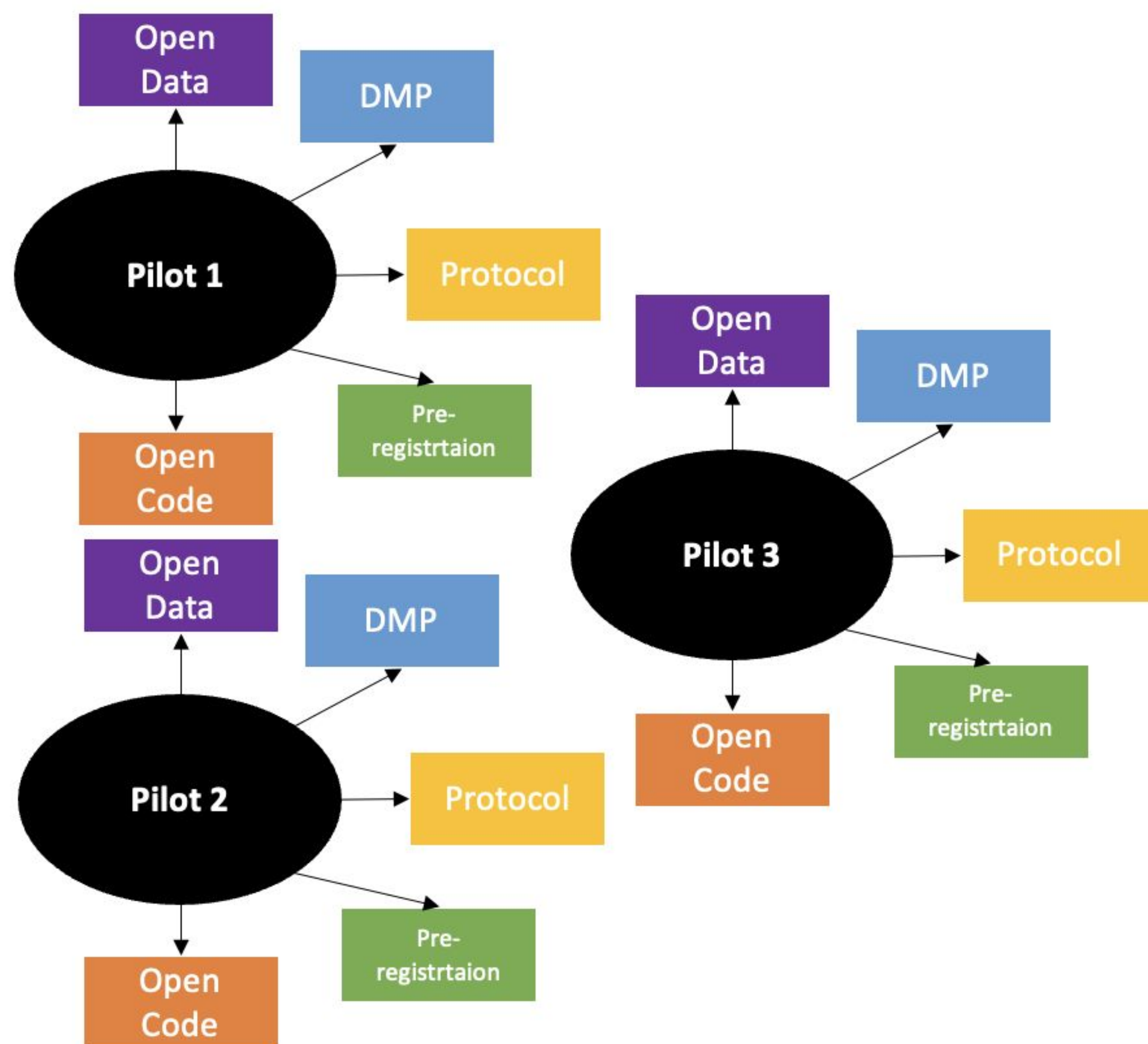
**Adoption &  
dissemination**



# Workflows



- Research workflow
  - Research design with FAIRness in mind
  - Keeping track of resources needed for implementing FAIR practices
- Metadata workflow
  - Generating rich metadata through DOI registration and system integrations
  - Bridging the generic to the specific through domain specific metadata templates



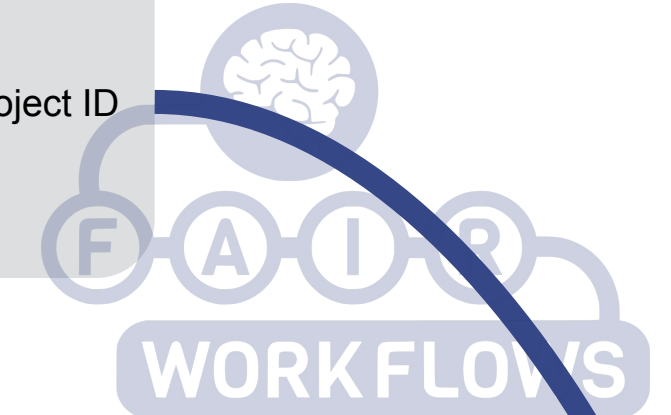


# FAIR Workflows Project

Potential workflow



Project ID



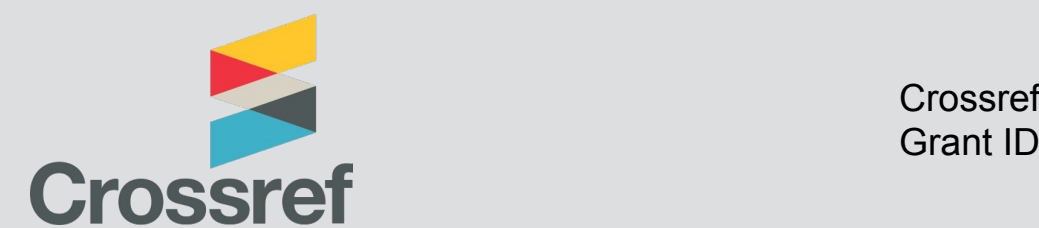
## 4. Research



## 3. Submit Data Management Plan



## 2. Application Approved



## 1. Grant Application



## 5. Protocols and research methods



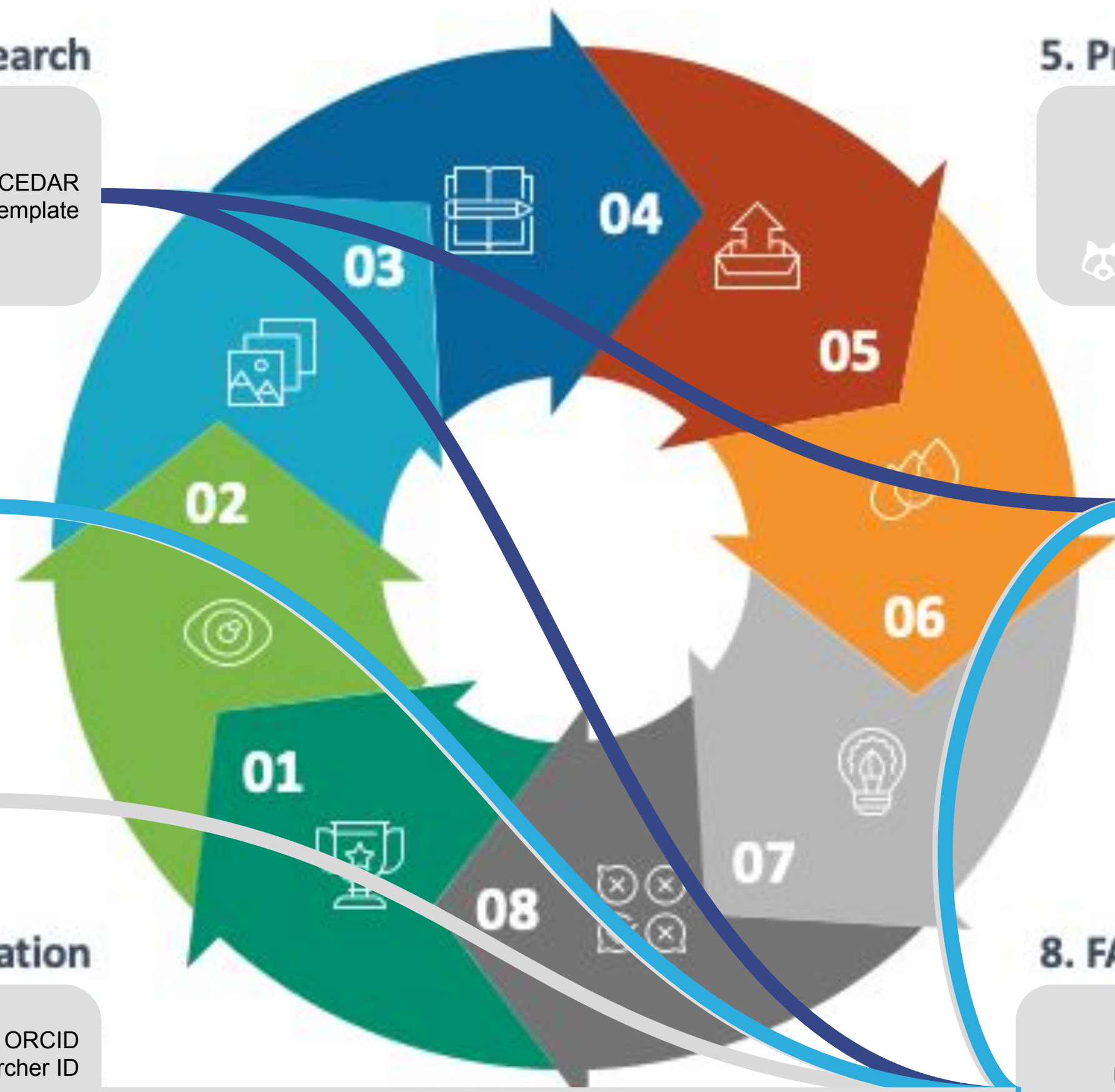
## 6. Research data, instruments deposited



## 7. Research article published



## 8. FAIR Workflow interface



# Dashboard



- Showcase the diversity of research outputs
  - Increase valuation of sharing beyond papers
- Indicate availability and connectedness
  - Facilitate reuse by lowering the barrier for discovery and acquisition
  - Enhance reproducibility by providing contextual information
- Track impact and credits
  - Support effective reporting and evaluation



Type to search...

Works

Go to Repository

Find Related Works

Contacts

curator@datadryad.org  
help@datadryad.org

Share

- Email
- Twitter
- Facebook

Type to search...

Works People Organizations

https://ror.org/03yrm5c26

California Digital Library

CDL

Links

Homepage

Wikipedia

Twitter

Geolocation

37° 48' 31.2415" N, 122° 16' 02.9173" E

United States Archive DataCite Consortium

https://ror.org/03yrm5c26

Share

- Email
- Twitter
- Facebook

Aggregated Citations, Views and Downloads

3 Citations

18 Works



Eurytemora affinis 2008 Experiment

Carly Strasser  
Version 1 of Dataset published 2012 in **DRYAD**  
Experiments to determine population growth rate  
DOI registered November 17, 2016 via DataCite.

CC BY

74 Views 4 Downloads

Dataset English

https://doi.org/10.15146/r3wk5c

EZID DOI Service is Evolving

Patricia Cruse, Michael Witt & Joan Starr  
Version 1.0 of Content published 2017 in **DataCite**  
The California Digital Library (CDL) and Purdue University (DOI) services to support DataCite's long-term sustainability.  
DOI registered August 5, 2017 via DataCite.

Text

https://doi.org/10.5438/0x88-gvge

Data: Researcher Perspectives on the

Yasmin Alnoamany & John Borghi  
Version 3 of Dataset published 2018 in **UC Berkeley**  
We are interested in learning about perceptions, values, and processes. To understand researchers' perspectives on data management at academic institutions throughout the United States, we conducted a survey of academic responses of the survey participants after excluding by the University of California Berkeley Committee

Type to search...

Works People Organizations Repositories

https://doi.org/10.5061/dryad.4hs71t6t

Data from: Mitochondrial DNA assessment of the phylogeography of the gopher tortoise

Joshua R. Ennen, Brian R. Kreiser, Carl P. Qualls, Daniel Gaillard, Matthew Aresco, Henry Mushinsky, Thomas W. Hentges & Aaron Schrey  
Version 1 of Dataset published 2012 in **DRYAD**  
Identifying geographic barriers that define genetic structure within a species is crucial for the protection of appropriate management units is critical for the protection of polyphemus, which have declined across their entire range. Previous molecular work has identified several potential geographic barriers. We sequenced a 712-base-pair portion of the mitochondrial cytb gene from 322 individuals from 42 sites across the range. We found two major assemblages of gopher tortoises (Tombigbee-Mobile). However, the presence of distinct (group 1 and 2) haplotypes indicates that the two lineages experienced historical isolation and divergence. We used these results to define genetic units of conservation for gopher tortoises, then the Apalachicola-Chattahoochee and Tombigbee-Mobile rivers as geographic barriers. We recommend that making management decisions on the basis of mitochondrial data should be based on the review additional work that examines finer scale patterns of genetic structure by using microsatellites.

DOI registered March 30, 2012 via DataCite.

CC BY

1 Citation 153 Views 75 Downloads

Dataset English

https://doi.org/10.5061/dryad.4hs71t6t

Creators

Joshua R. Ennen, Daniel Gaillard, Tracey D. Tuberville, Thomas W. Hentges, Brian R. Kreiser, Matthew Aresco, Earl D. McCoy, Aaron Schrey

Download

Full Metadata: DataCite XML, DataCite JSON, Schema.org JSON-LD, Citation Metadata: Citeproc JSON, BibTeX, RIS

ORCID Claim

Claim DOI

Share

Email, Twitter, Facebook

Cite

Ennen, J. R., Kreiser, B. R., Qualls, C. P., Gaillard, D., Aresco, M., Birkhead, R., Tuberville, T. D., & Schrey, A. (2012). Data from: Mitochondrial DNA assessment of the phylogeography of the gopher tortoise. *DRYAD*. <https://doi.org/10.5061/DRYAD.4HS71T6T>

153 Views 75 Downloads

153 views reported since publication in 2012.

1 Reference 1 Citation

Publication Year

Work Type

1

Filter Works

Type to search...

Publication Year

2012

Work Type

Text

Registration Agency

Crossref

Mitochondrial DNA Assessment of the Phylogeography of the Gopher Tortoise

Joshua R. Ennen, Brian R. Kreiser, Carl P. Qualls, Daniel Gaillard, Matthew Aresco, Henry Mushinsky, Thomas W. "Bill" Hentges & Aaron Schrey  
Journal Article published 2012 in **Journal of Fish and Wildlife Management**  
DOI registered via Crossref.  
1 Citation  
Journal Article  
https://doi.org/10.3996/102011-jfwm-063

Type to search...

Works People Organizations Repositories

https://orcid.org/0000-0001-6660-6214

Helena Cousijn

Helena Cousijn is DataCite's Community Engagement and Communications Director. Helena is responsible for all DataCite's outreach activities. She's committed to DataCite's mission of enabling data sharing and reuse and is especially passionate about data citation. Her main goal is to communicate in a way that makes DataCite's services accessible to everyone. Before joining DataCite, Helena worked as Senior Product Manager for Research Data Management Solutions at Elsevier. She holds a DPhil in Neuroscience from the University of Oxford.

Other Identifiers

GitHub: hcousijn

Other Profiles

ORCID

Impactstory

Europe PMC

Netherlands

https://orcid.org/0000-0001-6660-6214

Share

- Email
- Twitter
- Facebook

Employment

**DataCite**  
Director of Community Engagement  
Since July 2018

**Elsevier BV**  
Senior Product Manager  
March 2015 to June 2018

Aggregated Citations, Views and Downloads

15 Citations 11 Views

Accessibility Achievements

58% of the researcher's associated DOIs have metadata with rights as CC-BY, CC0 or public domain license.

104 Works



Presenting PIDapalooza 2019!

Helena Cousijn  
Version 1.0 of Content published 2018 in **DataCite Blog**  
PIDapalooza, the open festival of persistent identifiers is back and it's better than ever. Mark your calendar for Dublin, Ireland, January 23-24, 2019 - and send us your session ideas by September 21. Yes, it's back and - with your support - it's...

DOI registered August 28, 2018 via DataCite.

Text

https://doi.org/10.5438/cd2b-xj80

Taking discoverability to the next level: datasets with DataCite DOIs can now be found through

Google Dataset Search  
Helena Cousijn, Patricia Cruse & Martin Fenner

humanities and social sciences

bioinformatics and computational biology

biochemistry and molecular biology

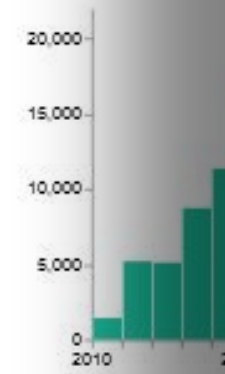
agriculture, forestry and land use

natural sciences (other)

biodiversity

148,840 Deposits

Year of Deposit



Field of Science



If you plan to deposit your data

More information

- Borghi, J., Alnoamany, Y., & Fenner, M. (2018). Researcher Perspectives on Data Management at Academic Institutions. *Journal of Fish and Wildlife Management*.
- Goodman, J., & Fenner, M. (2018). Care and Feeding of Data: A Survey of Academic Researchers. *Journal of Fish and Wildlife Management*.
- Pampel, H., & Fenner, M. (2018). Repositories and Data Management: A Survey of Academic Researchers. *Journal of Fish and Wildlife Management*.



# Adoption and dissemination



- Keep the communities informed and involved
  - Engaging both the neuroscience community and the open science community
  - Build synergy, avoid reinventing the wheel
- Provide open source code for the dashboard
- Build best practices guide for researchers
- Produce technical documentation and metadata guidelines for future integrators



**The project team will reach out and seek feedback from the community to help us test the workflows and dashboard in different contexts. Directly email me to be added to the Friends of the Project list:**

**New Message**

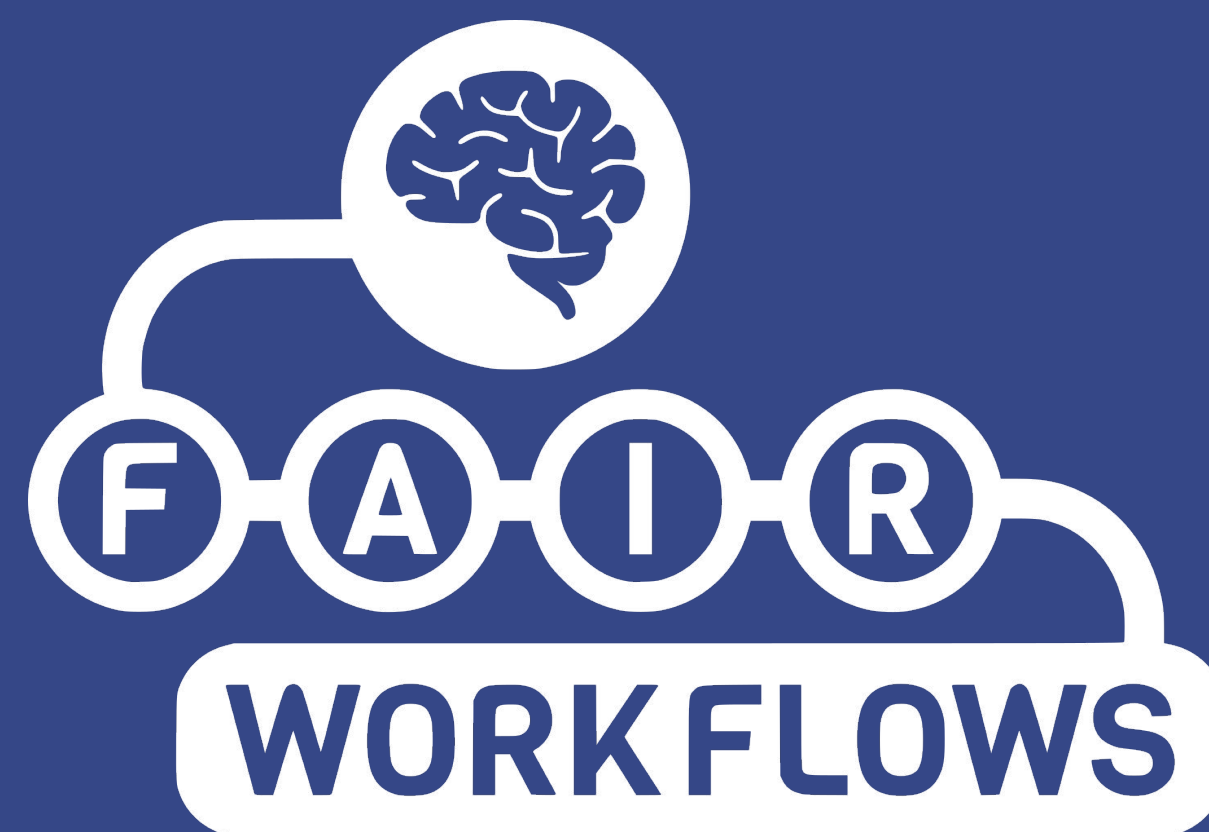
To

[xiaoli.chen@datacite.org](mailto:xiaoli.chen@datacite.org)

Cc Bcc

Subject

**[your org.] - count me in with FAIR Workflows**



**Keep in touch!**



[xiaoli.chen@datacite.org](mailto:xiaoli.chen@datacite.org)



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



[slack](#)



<http://datacite.org/fair-workflows.html>



... it is our intent that the principles apply not only to 'data' in the conventional sense, but also to the algorithms, tools, and workflows that led to that data.

All scholarly digital research objects—from data to analytical pipelines—benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability.