

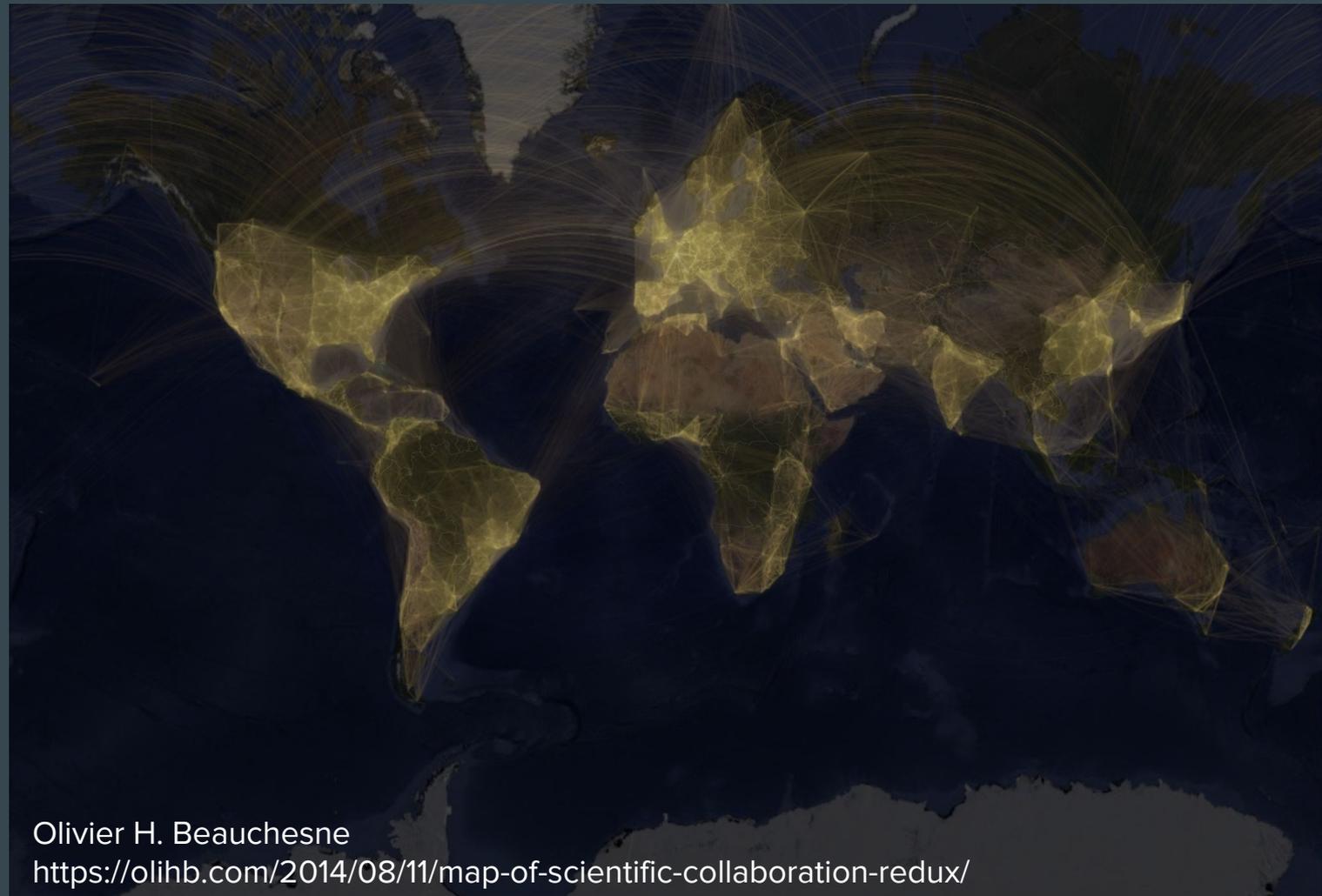
Increasing the yield of research through openness

Carly Strasser | @carlystrasser

Crops in silico 2018 | UIUC



Research has changed



Olivier H. Beauchesne

<https://olihb.com/2014/08/11/map-of-scientific-collaboration-redux/>



CC-BY Andrew Taylor



docker



python



binder



Trust in

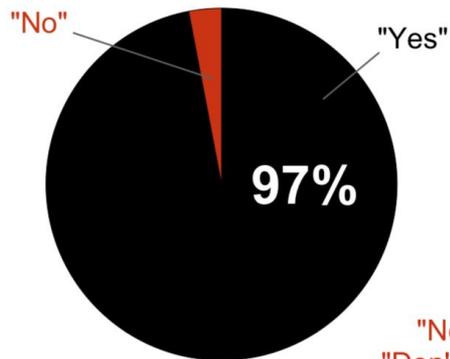
Research has changed

San Diego Historical Society

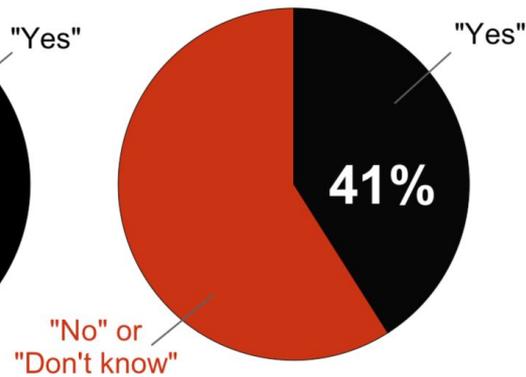


Say Climate Change is Happening and Human Caused

Climate Scientists



American Public



Yale Project on Climate Change Communication



Over half of psychology replication studies fail

Largest replication study to date

Monya Baker

27 August 2015

THINK PSYCHOLOGY'S REPLICATION CRISIS IS BAD? WELCOME TO THE ONE IN MEDICINE.



SL

FUTURE TENSE | THE CITIZEN'S GUIDE TO THE FUTURE. | APRIL 19 2016 9:21 AM

FROM SLATE, NEW AMERICA, AND ASU

Cancer Research Is

Science

Home News Journals Topics Careers



Log in | My account | C

SHARE IN DEPTH | COMPUTER SCIENCE



0



0

Artificial intelligence faces reproducibility crisis

Matthew Hutson

+ See all authors and affiliations

Science 16 Feb 2018:
Vol. 359, Issue 6377, pp. 725-726
DOI: 10.1126/science.359.6377.725

Article

Figures & Data

Info & Metrics

eLetters

PDF

Summary

The booming field of artificial intelligence (AI) is grappling with a replication crisis, much like the ones



THE CONVERSATION US PILOT

Academic rigor, journalistic flair

[Arts + Culture](#) [Economy + Business](#) [Education](#) [Environment + Energy](#) [Health + Medicine](#) [Politics + Society](#) [Science + Technology](#) [Election 2016](#)

How do we solve science's 'credibility problem'?

April 7, 2015 9:42am EDT

chuttersnap/Unsplash

Been

The Washington Post

ore

The Fix

Americans' increasing distrust of science — and not just on climate change

A



266



Save for Later



Reading List

Trust in
Research has changed

...and this affects funding!

SHARE



SHARE
240



TWEET



COMMENT



EMAIL

MICHAEL SCHULSON SCIENCE 04.20.18 09:00 AM

SCIENCE'S "REPRODUCIBILITY CRISIS" IS BEING USED AS POLITICAL AMMUNITION



CHEMISTRY WORLD

CHEMISTRY

Battle over Science Funding Gets Fiercer in U.S. Congress

Research agency critics, such as Sen. Ted Cruz, have new positions with power to control money and priorities

NEW REPUBLIC

Mark Wilson / Getty

Republicans' War on Science Just Got Frighteningly Real

Two terrible bills that Obama would have vetoed are back in play, and they stand a very good chance of becoming law under Trump.

By **EMILY ATKIN** | March 9, 2017

As many conservatives see it, environmental science is an enabler of dreaded

Open Science

What can we do?

Open Science:

Making **process**
outputs

of research available to all

publications
datasets
images
software/code
proposals
reviews
other

outputs

thought process
workflows
code
output generation

process

Open Science

Enables

reproducibility

credibility

reuse

participation

collaboration!

publications

datasets

images

software/code

proposals

reviews

other

outputs

Preprints & Open Access

Preprint: a manuscript that has not yet been peer reviewed for formal publication. These may be drafts, incomplete versions, or final versions of articles.

Benefits

- Faster dissemination
- Establish precedent
- Solicit feedback
- Open access

The logo for ASAPbio, featuring a red chevron pointing right followed by the text "ASAPbio" in black.The logo for PeerJ, with the word "PeerJ" in blue.The logo for bioRxiv, with the text "bio" in large black letters and "THE PREPRINT" in smaller black letters below it.

Preprints & Open Access

No price or permission barriers

Full content

Immediately available

Not just publications



in silico
Plants

Introducing
in silico Plants

in silico Plants (isP) is an innovative new
open-access journal dealing with all aspects of plant
modelling.

Guest Editor
Professor Stephen Long

Visit academic.oup.com/insilicoplants

for more information, instructions to authors, and to submit online today.

OXFORD
UNIVERSITY PRESS

Introducing
in silico Plants

in silico Plants (isP) is an innovative new
open-access journal dealing with all aspects of
modelling.

Yay!!

publications

datasets

images

software/code

proposals

reviews

other

outputs



From Flickr by Ninja M.

Open data

certain data should be **freely available** to everyone to **use & republish** as they wish, **without restrictions** from copyright, patents or other mechanisms of control



From Flickr by Ninja M.

Open data

...is only useful with
high-quality metadata!

Data management 101

- Data collection & organization
- Short term storage and backing up
- Quality control & assurance
- Metadata generation
- Workflow documentation
- Long term storage and preservation plans

publications

datasets

images

software/code

proposals

reviews

other

outputs

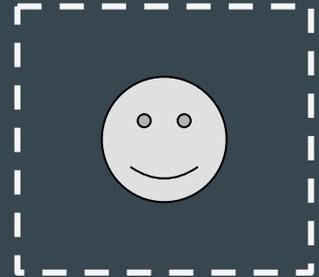
Open Source Software

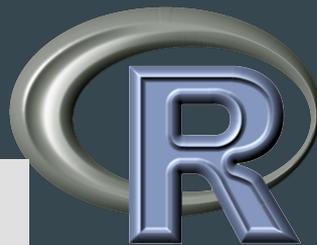
- sou
 - lice
 - sof
 - De
- Benefits**
- Reproduce research
 - Find & fix bugs
 - Extend utility: new software
 - Community effort
- change and distribute the
ose
manner



Open Source Software

- source code available
- licensed so that others can study, change and distribute the software to anyone and for any purpose
- Developed in a public, collaborative manner





Benefits

- Reproduce research
- Find & fix bugs
- Extend utility: new software
- Community effort

The Julia logo, featuring the word "julia" in white lowercase letters with three white dots above the "i".

julia

17 December 2014

Software

Open access

Researcher Data Publication Perceptions

[John Ernest Kratz](#) ; [Carly Strasser](#)

(show affiliations)

Data "publication" seeks to appropriate the prestige of authorship in the peer-reviewed literature to reward researchers who create useful and well-documented datasets. The scholarly communication community has embraced data publication as an incentive to document and share data. But, numerous new and ongoing experiments in implementation have not yet resolved what a data publication should be, when data should be peer-reviewed, or how data peer review should work. While researchers have been surveyed extensively regarding data management and sharing, their perceptions and expectations of data publication are largely unknown. To bring this important yet neglected perspective into the conversation, we surveyed 249 researchers across the sciences and social sciences— asking what expectations "data publication" raises and what features would be useful to evaluate the trustworthiness, evaluate the impact, and enhance the prestige of a data publication. We found that researcher expectations of data publication center on availability, generally through an open database or repository. Few respondents expected published data to be peer-reviewed, but peer-reviewed data enjoyed much greater trust and prestige. The importance of adequate metadata was acknowledged, in that almost all respondents expected data peer review to include evaluation of the data's documentation. Formal citation in the reference list was affirmed by most respondents as the proper way to credit dataset creators. Citation count was viewed as the

Available in



Publication date:

17 December 2014

DOI

DOI [10.5281/zenodo.13293](https://doi.org/10.5281/zenodo.13293)

Keyword(s):

[Data publication](#) [Data](#)

Related publications and datasets:

Supplement to:

<https://github.com/JEK-III/ResearcherDataPubPerceptions/tree/v1.0>

Collections:

[Software](#)[Open Access](#)

License (for files):

mybinder.org



Turn a GitHub repo into a collection of
interactive notebooks

Have a repository full of Jupyter notebooks? With Binder, open those notebooks in an executable environment, making your code immediately reproducible by anyone, anywhere.

Sharing and Publishing Jupyter Notebooks

Documentation

🔍 Overview

Teaching: 10 min

Exercises: 15 min

Questions

- Why should I invest time in good documentation?
- How does my target audience influence my documentation strategy?
- What are some published examples of good documentation?

Objectives

- Describe how documentation is useful to yourself and to others
- Evaluate and rank the quality of comments in published notebooks
- Evaluate and rank the quality of existing metadata records.
- Describe types of metadata directly relevant for research reproducibility.

Overview

Documenting your process, especially as it concerns your data, is a key element of making your research more reproducible. If you do not thoroughly record all the data manipulation steps you used to process data, it will likely be impossible for you, or anyone else, to repeat the analysis in the future (Wilson et al. 2016). Using the Jupyter Notebook for scripting your data processing is powerful because it saves the code – the **what** – and interspersed it the motivations behind each step, i.e., the **why**.

There is also project-level documentation that isn't needed to understand a particular series of data processing steps, but to understand the organization of the project as a whole. Finally, documentation can be used to aid discoverability.



From Flickr by Ninja M.

Open *software*

...is only useful with
high-quality *documentation*



DATA CARPENTRY

MAKING DATA SCIENCE MORE EFFICIENT

Data Carpentry develops and teaches workshops on the fundamental data skills needed to conduct research. Our mission is to provide researchers high-quality, domain-specific training covering the full lifecycle of data-driven research. Data Carpentry is a sibling organization of Software Carpentry. Where Software Carpentry teaches best practices in software development, our focus is on the introductory computational skills needed for data management and analysis in all domains of research. Our lessons are domain specific, from life and physical sciences to social science and build on the existing knowledge of learners to enable them to quickly apply skills learned to their own research. *Our initial*

christ@illinois.edu
christ@illinois.edu
christ@illinois.edu
christ@illinois.edu
christ@illinois.edu

Code is Science Manifesto

- 🔥 Open over closed
- 🔥 Incorrect code results in incorrect science
- 🔥 Code for the future
- 🔥 Availability over perfection
- 🔥 Code deserves credit

Scientific code needs to be treated
as a primary research output



**Code (is)
Science**



Pledge support (or volunteer!) at
codeisscience.com/manifesto

ay.

rights?
assment,
o call

publications

datasets

images

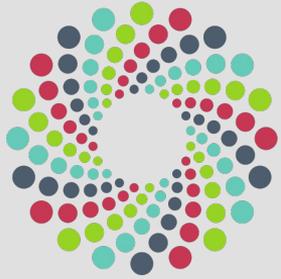
software/code

proposals

reviews

other

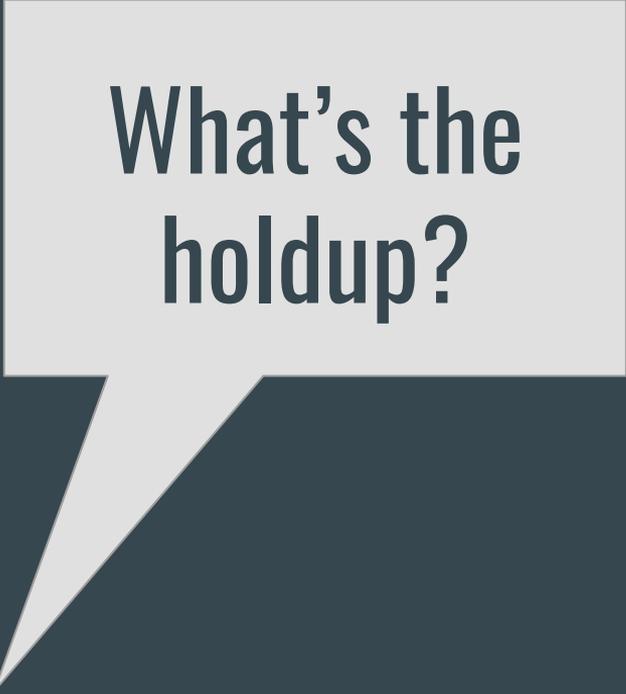
outputs



figshare

INSTITUTIONAL
REPOSITORIES

zenodo



**What's the
holdup?**

- 
- A long, perspective view of a track with hurdles, with a semi-transparent text box overlaid in the center. The hurdles are blue and white, with the words "GILL EAGLES GILL" printed on them. The track is dark with white lane markings. In the background, several people are visible, some standing and some walking. The scene is lit with warm, golden light, suggesting late afternoon or early morning.
- Cost
 - Lack of training
 - Fear of lost rights or benefits
 - No incentives

From Flickr by Gord Fynes



A dense collage of various newspaper clippings, advertisements, and posters is mounted on a dark, textured background. The collage includes numerous environmental and nature-themed items, such as posters for 'NATURE Oakland', 'PARK HERE ANY TIME', and 'SOUTH ALBANY'. There are also several posters featuring ducks and other wildlife. The text 'Articles are only the advertisements for scholarship.' is overlaid in white on a semi-transparent dark rectangular background in the center of the collage.

Articles are only the advertisements
for scholarship.

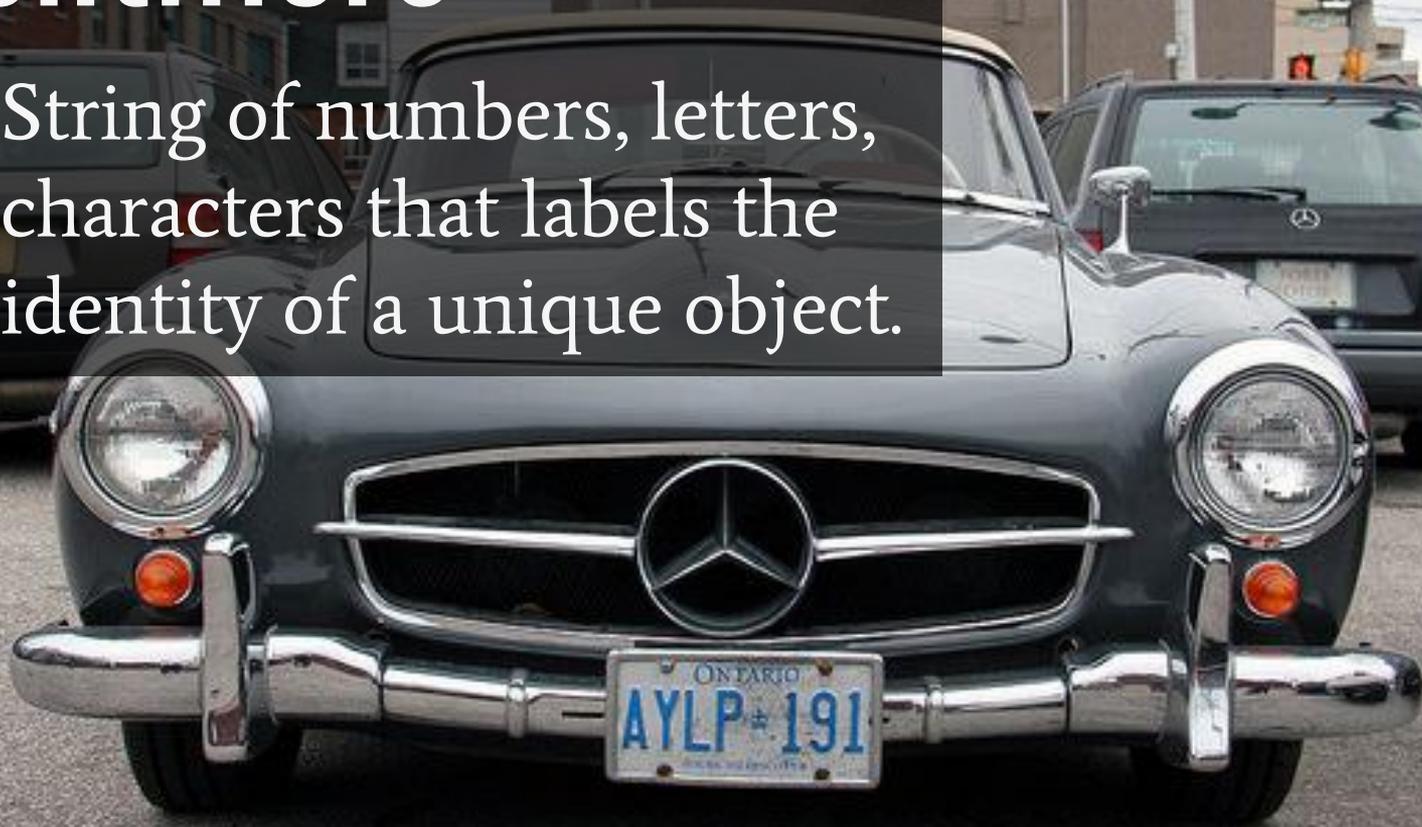
How to make **process**
outputs count?

1. Share them
2. Connect them
3. Claim them
4. Advertise them

get persistent
identifiers &
cite them

Identifiers

String of numbers, letters, characters that labels the identity of a unique object.



Identifiers in various disciplines [\[edit\]](#)

A small sample of [various identifiers](#)

Identifier	Scope
atomic number, corresponding one-to-one with element name	international (via ISV)
Australian Business Number	Australian
CAGE code	U.S. and NATO
CAS registry number	originated in U.S.; today international (via ISV)
CO	
Dig	pe
DIN	
E number	originated in E.U.; may be seen internationally
EC number	
Employer Identification Number (EIN)	U.S.
Global Trade Item Number	international
Group identifier	many scopes, e.g., specific computer systems
International Chemical Identifier	international
International Standard Book Number (ISBN)	ISBN is part of the EAN Namespace ; international scope
International Standard Serial Number (ISSN)	international
ISO standard number, e.g., ISO 8601	international
Library of Congress Control Number	U.S., with some international bibliographic usefulness
Personal identification number	many scopes, e.g., banks, governments
Personal identification number (Denmark)	Denmark

Digital Object Identifier, or DOI

Identifiers are important for citation!



data

software

genetic code

reagents

protocols

...

Data citation

Sidlauskas, B. 2007. Data from: Testing for unequal rates of morphological diversification in the absence of a detailed phylogeny: a case study from characiform fishes. Dryad Digital Repository. [doi:10.5061/dryad.20](https://doi.org/10.5061/dryad.20)

Allows readers to find data products

Get credit for data **and** publications

Promotes reproducibility



[Logout](#) | [Contact UC3](#) | [Help](#)

Logged in as Guest



Login

Object: [ark:/b5060/d8h59d](http://n2t.net/ark:/b5060/d8h59d)

Merritt > Collection: [UCOP Dash](#) > Object: [ark:/b5060/d8h59d](http://n2t.net/ark:/b5060/d8h59d)

Login was successful

Kratz, John; Strasser, Carly. **Making Data Count survey responses** (2015)

permanent link: <http://n2t.net/ark:/b5060/d8h59d>

primary identifier: [ark:/b5060/d8h59d](http://n2t.net/ark:/b5060/d8h59d)

title: Making Data Count survey responses

creator: Kratz, John; Strasser, Carly

date: 2015

local id: tcwfmlmipj; doi:10.5060/D8H59D

last modified: 2015-09-25 05:29 AM UTC

created: 2015-03-25 03:06 PM UTC

total size (all versions): 2.1 MB

total versions: 3

[Download object](#)

Current Version

Version 3: 2015-09-25 05:28 AM UTC

- [MDC_Managers.tsv](#)
- [MDC_Researchers.tsv](#)
- [MDC_Survey.pdf](#)
- [mrt-datacite.xml](#)

[3 more files](#)

Prior Versions

Version 2: 2015-07-02 09:48 AM UTC

Version 1: 2015-03-25 03:06 PM UTC

Merritt is a service of the [University of California Curation Center](#) of the [California Digital Library](#)

Copyright © 2009-2016 The Regents of the University of California

[Privacy policy](#) | [Image credits](#)

17 December 2014

Software

Open access

Researcher Data Publication Perceptions

[John Ernest Kratz](#) ; [Carly Strasser](#)

(show affiliations)

Data "publication" seeks to appropriate the prestige of authorship in the peer-reviewed literature to reward researchers who create useful and well-documented datasets. The scholarly communication community has embraced data publication as an incentive to document and share data. But, numerous new and ongoing experiments in implementation have not yet resolved what a data publication should be, when data should be peer-reviewed, or how data peer review should work. While researchers have been surveyed extensively regarding data management and sharing, their perceptions and expectations of data publication are largely unknown. To bring this important yet neglected perspective into the conversation, we surveyed 249 researchers across the sciences and social sciences— asking what expectations "data publication" raises and what features would be useful to evaluate the trustworthiness, evaluate the impact, and enhance the prestige of a data publication. We found that researcher expectations of data publication center on availability, generally through an open database or repository. Few respondents expected published data to be peer-reviewed, but peer-reviewed data enjoyed much greater trust and prestige. The importance of adequate metadata was acknowledged, in that almost all respondents expected data peer review to include evaluation of the data's documentation. Formal citation in the reference list was affirmed by most respondents as the proper way to credit dataset creators. Citation count was viewed as the

Available in



Publication date:

DOI

DOI [10.5281/zenodo.13293](https://doi.org/10.5281/zenodo.13293)

Keyword(s):

[Data publication](#) [Data](#)

Related publications and datasets:

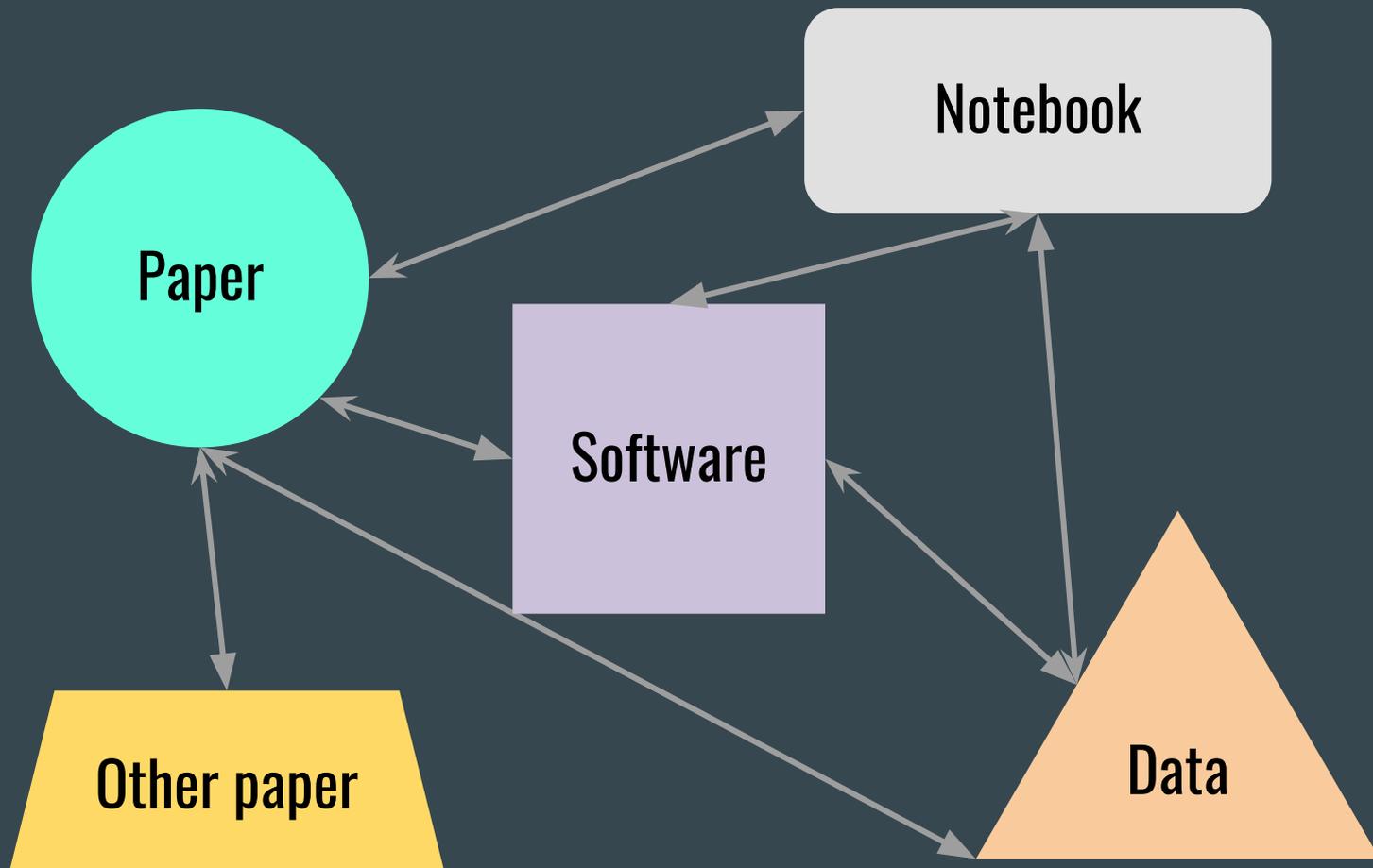
Supplement to:

<https://github.com/JEK-III/ResearcherDataPubPerceptions/tree/v1.0>

Collections:

[Software](#)[Open Access](#)

License (for files):



A photograph of a crowd at a concert or event. In the foreground, several pairs of hands are raised, forming a large heart shape. The background is filled with people, some of whom are also forming heart shapes with their hands. The lighting is warm and golden, suggesting a sunset or stage lights. The overall mood is celebratory and communal.

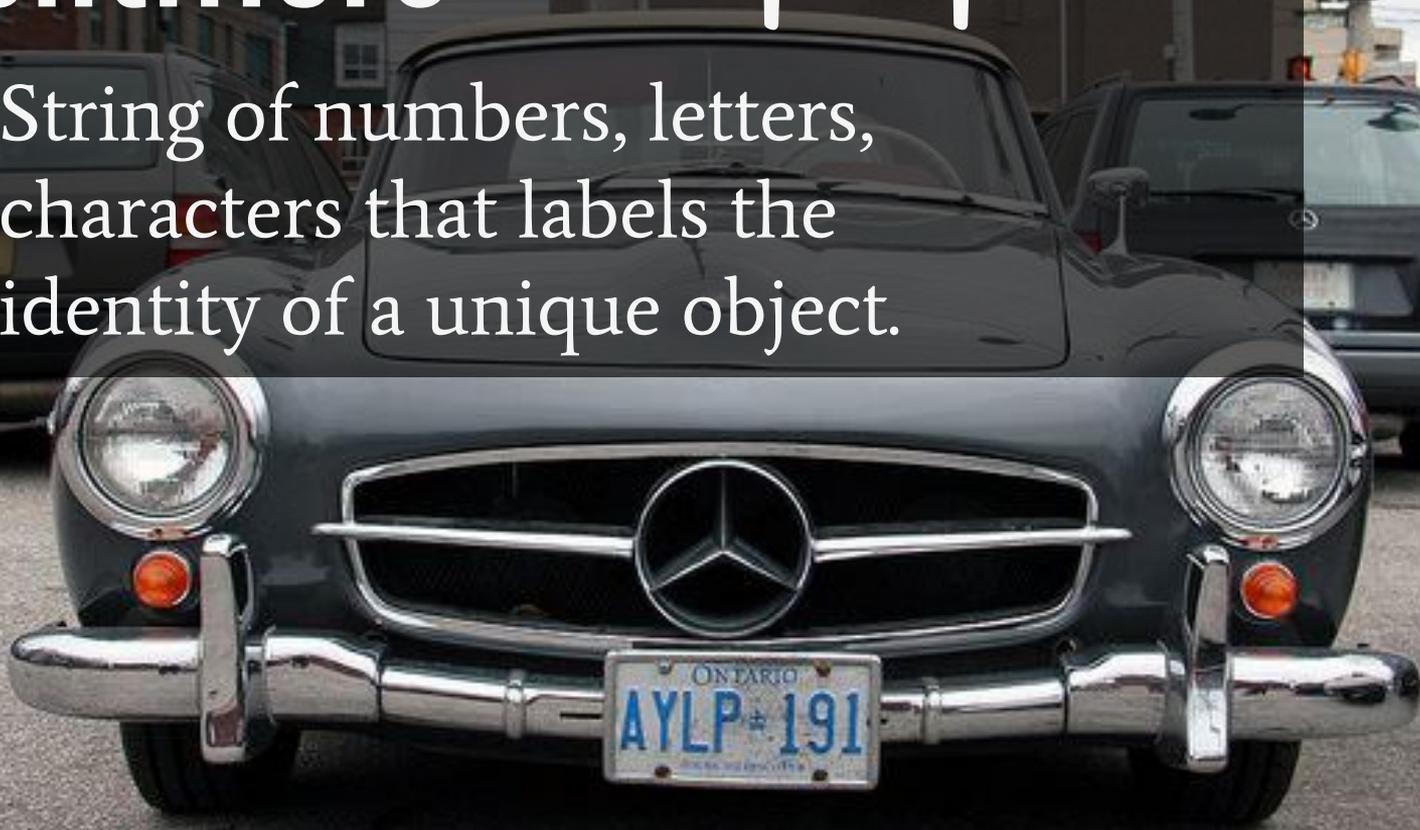
PIDapalooza

SOLD OUT!

1. Share them
2. Connect them
3. Claim them
4. Advertise them

Identifiers For people

String of numbers, letters, characters that labels the identity of a unique object.



DISTINGUISH YOURSELF IN THREE EASY STEPS

ORCID provides a persistent digital identifier that distinguishes you from every other researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between you and your professional activities ensuring that your work is recognized. [Find out more.](#)

1

REGISTER Get your unique ORCID identifier [Register now!](#)
Registration takes 30 seconds.

2

**ADD YOUR
INFO** Enhance your ORCID record with your
professional information and link to your other
identifiers (such as Scopus or ResearcherID or
LinkedIn).

3

USE YOUR Include your ORCID identifier on your Webpage,
when you submit publications, apply for grants, and



LATEST NEWS

Tue, 2016-05-31
[Collect & Connect:
Turning ORCID's
Vision into Reality](#)

Thu, 2016-05-19
[What was my
password again?](#)

Funders



AAAS
AGU
eLife
EMBO
Hindawi
IEEE
PLOS
Royal Society



Publishers



Carly Strasser

Fred Hutchinson Cancer Research Center Director of Academic Alliances & Data Strategy

 share

OVERVIEW

ACHIEVEMENTS

TIMELINE

PUBLICATIONS

ACHIEVEMENTS

[view all](#)



Open Access Top 10%

97% of your research is free to read online. This level of availability puts you in the top 1% of researchers.



Global Reach Top 10%

Your research has been saved and shared in 56 countries. That's high: only 4% of researchers get that much international attention.



Greatest Hit Top 10%

Your top publication has been saved

TIMELINE

[view all](#)

1806 Online mentions
over 1 years

 1.7k  51  38  21  12  2  2

PUBLICATIONS

[view all](#)

 [Preprints for the life sciences](#)

2016 *Science*

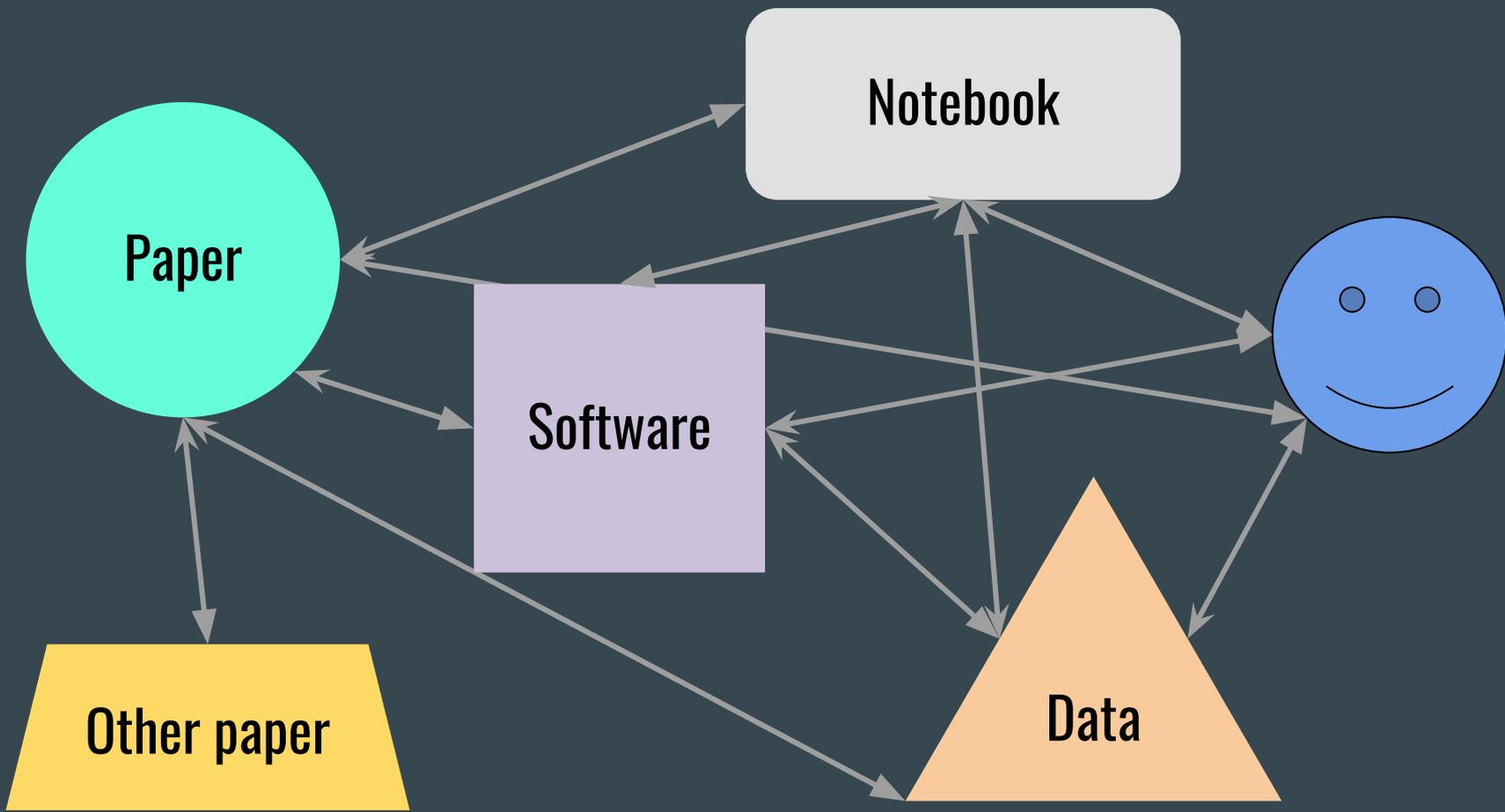
592     

 [Making data count](#)

2015 *Scientific Data*

258     

 [The Role of Publishers in Access to Data](#)



1. Share them
2. Connect them
3. Claim them
4. Advertise them

Carly A. Strasser

CONTACT INFORMATION

EDUCATION

RESEARCH & PROFESSIONAL EXPERIENCE

OTHER RESEARCH OUTPUTS

C Strasser and E Khare. 2017. **Data**. Moore grantee list of articles. Zenodo. doi:[10.5281/zenodo.841794](https://doi.org/10.5281/zenodo.841794). & Moore grantee publication data. Zenodo. doi:[10.5281/zenodo.555947](https://doi.org/10.5281/zenodo.555947).

C Strasser, P Cruse, J Kunze, and S Abrams. 2013. **Data** from “DataUp: A tool to help researchers describe and share tabular data”. figshare. doi: [10.6084/m9.figshare.884625](https://doi.org/10.6084/m9.figshare.884625)

CA Strasser and SE Hampton. 2012. **Data** from “The Fractured Lab Notebook: Undergraduates and Ecological Data Management Training in the United States.” [Data in KNB Repository](#).

C Strasser, P Cruse, and S Abrams. 2014. **Grant** from the Alfred P. Sloan Foundation. “Dash: Improving Community Repositories for Better Data Sharing.” Available at escholarship.org/uc/item/2mw6v93b.

J Lin, P Cruse, M Fenner, and C Strasser. 2014. **Grant** from the National Science Foundation, EAGER ACI Award No. 1448821. “Making Data Count: Developing a Data Metrics Pilot.” Available at escholarship.org/uc/item/9kf081vf.

P Cruse and C Strasser. 2012. **Grant** from Institute of Museum and Library Services, No. LG-07-12-0505-12. “Improving Data Stewardship with the DMPTool: Empowering Libraries to Seize Data Management Education.”

P Cruse, S Abrams, C Strasser, A Sallans, S Shreeves, and A Budden. 2012. **Grant** from the Alfred P. Sloan Foundation. “Data Management Planning Tool 2: Responding to the Community.”

P Cruse, C Strasser, W Michener, J Kunze, and D Vieglais. 2012. **Grant** from the National Science Foundation, supplemental to OCI Award No. 0830944. “DataUp: Further Development and Community Building.” Available at escholarship.org/uc/item/1dt1v1pc.

High quality
Trustworthy
Collaborative
Open
Transparent
Reproducible
Reusable

Science

website

carlystrasser.net

email

carlystrasser@gmail.com

twitter

@carlystrasser



FRED HUTCH
CURES START HERE™