Data sharing[1]
Why, when, where, & how

Rick Gilmore   Bruce Desmarais

Psychology/CSC

Political Science/SoDA

Preliminaries

Agenda

Why share data

Transparency

Accelerate discovery

Improve rigor

Enough already. Five different teams of physicists have now independently verified that elusive subatomic particles called neutrinos do *not* travel faster than light. New results, announced today in Japan, contradict those announced last September by a 170-member crew working with the OPERA particle detector in Italy's subterranean Gran Sasso National Laboratory. The OPERA team made headlines after they suggested neutrinos traveled 0.002% faster than light, thus violating Einstein's theory of special relativity. The OPERA results were debunked months ago, however. So instead of the nail in the coffin of faster-than-light neutrinos, the new suite of results is more like the sod planted atop their grave.

Internal reproducibility

Could someone else on your team reproduce analyses done by X? -
Methods reproducibility, [**?**]

Requirement
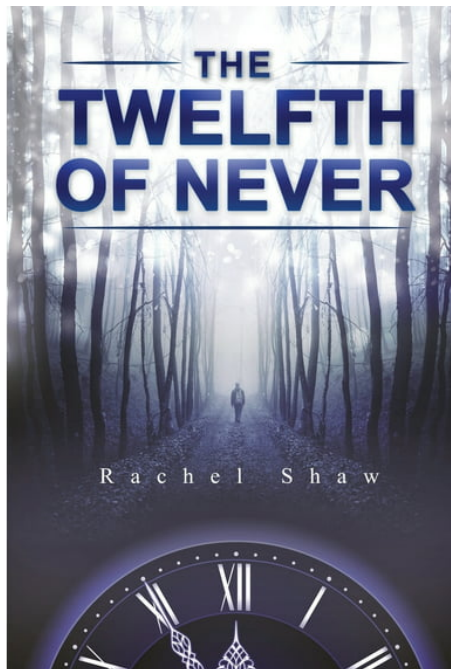
- Funder
- Journal
- Institutional

Ethical duty

Whereas regulators of human subjects research often view data sharing solely in terms of potential risks to subjects, we argue that the principles of human subject research require an analysis of both risks and benefits, and that such an analysis suggests that **researchers may have a positive duty to share data** *in order to maximize the contribution that individual participants have made.*

[**?**]

When to share

When NOT to share…

▶ When you've milked every last ounce of information out of a dataset, and all that remains is a lifeless, shriveled husk

On submission

▶ Some journals require that authors submit data for review. Ex, *Nature Scientific Data*

2 Select a repository for your data ⬆

When submitting a **Data Descriptor**, authors must deposit all relevant datasets in an appropriate public repository prior submission, and the completeness of these datasets will be considered during editorial evaluation and peer-review. Datasets must be made publicly available without restriction in the event that the Data Descriptor is accepted for publication (except reasonable controls related to human privacy issues or public safety).

- Reviewers have the option to review the data along with the manuscript. - Bruce had experience of reviewer at NSD pointing out that numbers from the manuscript did not match data repository.

After acceptance

▶ Pre-publication verification increasingly common

On publication

End of grant period

Where to share data

Data repository

- Dataverse (linked to journal article)
- Databrary.org (esp. video/audio, identifiable data)
- ICPSR
- OSF.io

- PSU ScholarSphere
- PSU Data Commons
- NIH Data Archive
- Zenodo.org

Alternatives to repositories

▶ Supplemental materials
▶ Lab/project website

Data journal

List of data journals

🆔 Kindling, Maxi; 🆔 Strecker, Dorothea

This document describes a dataset that aggregates information about 135 data journals.
Data journals focus on the publication of data papers -- a specialized publication type describing datasets, their collection and reuse potential that is peer-reviewed, citable and indexed.
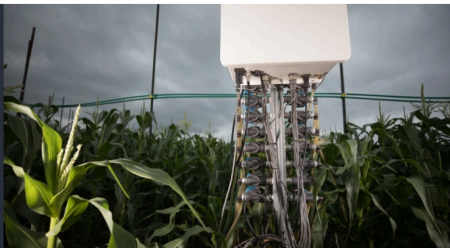This dataset includes a comprehensive list of data journals that was compiled by aggregating existing sources, as well as an overview of these sources.

The list is continually updated on GitHub, where additional information on data journals (URLs of data journal homepages) is provided: https://github.com/MaxiKi/data-journals

Preview			∨
issn	**journal_title**	**publisher**	**data_journal_type**
2363-4952	RIDE - A review journal for digital editions and resources	Institut für Dokumentologie und Editorik	pure
2059-481X	Journal of Open Humanities Data	Ubiquity Press	pure
2296-7745	Frontiers in Marine Science	Frontiers	mixed
2054-345X	Human Genome Variation	Springer Nature	mixed
2032-6378	Journal of Astronomical Data	Vrije Universiteit Brussel	pure
2603-431X	Viticulture Data Journal	Pensoft Publishers	pure

nature > scientific data

Two decades of fumigation data from the Soybean Free Air Concentration Enrichment facility

Elise Kole Aspray, Timothy A. Mies ... Elizabeth A. Ainsworth

Data Descriptor | 20 April 2023

Announcements

Collection open for submissions

Scientific Data is open to submissions for this special collection: Meteorology and hydroclimate observations and models

Open for submissions

www.nature.com/sdata/

n.b.– Gilmore is on editorial board.

- Could be extra pub
- Especially good for early-career researchers

How to share data

SINCE EVERYONE SENDS STUFF THIS
WAY ANYWAY, WE SHOULD JUST
FORMALIZE IT AS A STANDARD.

FAIR principles

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**esusable

[**?**]

CARE principles

- **C**ollective benefit
- **A**uthority to control
- **R**espect
- **E**thical use

[**?**]

Seek permission to share

▶ Sensitive & identifiable data can be shared with *restricted audiences* (i.e., researchers)

▶ e.g., Databrary

Resources

About

This talk was produced using Quarto. Rick Gilmore used the RStudio Integrated Development Environment (IDE). Bruce Desmarais used XXX.

The source files are in R and R Markdown, then rendered to PDF slides in the beamer format. The PDF slides are hosted in a GitHub repo:

https://github.com/penn-state-open-science/bootcamp-2023-data-sharing

References