

All's FAIR in love and publishing: *Science's* data journey, 2015-21

JAKE S. YESTON
EDITOR, *SCIENCE*

Science |  AAAS

Step 1: Formulating TOP (June 2015)

POLICY FORUM | SCIENTIFIC STANDARDS

Promoting an open research culture

B. A. Nosek^{*}, G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, T. Yarkoni

Affiliations for the authors, all of whom are members of the TOP Guidelines Committee, are given in the supplementary materials.

↩^{*}Corresponding author. E-mail: nosek@virginia.edu

- Hide authors and affiliations

Science 26 Jun 2015:
Vol. 348, Issue 6242, pp. 1422-1425
DOI: 10.1126/science.aab2374

Step 1: Formulating TOP (June 2015)

Summary of the eight standards and three levels of the TOP guidelines

Levels 1 to 3 are increasingly stringent for each standard. Level 0 offers a comparison that does not meet the standard.

	LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3
Citation standards	Journal encourages citation of data, code, and materials—or says nothing.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used, consistent with journal's author guidelines.	Article is not published until appropriate citation for data and materials is provided that follows journal's author guidelines.
Data transparency	Journal encourages data sharing—or says nothing.	Article states whether data are available and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Analytic methods (code) transparency	Journal encourages code sharing—or says nothing.	Article states whether code is available and, if so, where to access them.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.

Step 2: Adopting TOP Level 2 (January 2017)

EDITORIAL

Taking up TOP

Nearly 1 year ago, a group of researchers boldly suggested that the standards for research quality, transparency, and trustworthiness could be improved if journals banded together to adopt eight standards called TOP (Transparency and Openness Promotion).^{*} Since that time, more than 500 journals have been working toward their implementation of TOP. The editors at *Science* have held additional retreats and workshops to determine how best to adapt TOP to a general science journal and are now ready to announce our new standards, effective 1 January 2017.[†]



practice increases transparency, enables reproducibility, promotes data reuse, and is increasingly in line with funder mandates. We are also requiring the citation of all data, program code, and other methods not contained in the paper, using DOIs (digital object identifiers), journal citations, or other persistent identifiers, for the same reason. Citations reward those who originated the data, samples, or code and deposited them

for reuse. Such a policy also allows accurate accounting for exactly which specific data, samples, or code were used in a given study.

These guidelines also apply to our open-access journal *Science Advances*.



Marcia McNutt
Editor-in-Chief
Science Journals

Step 2: Adopting TOP Level 2 (January 2017)

TOP Guidelines

The *Science* Journals support the Transparency and Openness Promotion (TOP) **guidelines** to raise the quality of research published in *Science* and to increase transparency regarding the evidence on which conclusions are based.

1. Citation standards

All data, program code, and other methods must be appropriately cited using digital object identifiers (**DOIs**), journal citations, or other persistent identifiers. Exceptional circumstances requiring special treatment, such as proprietary information requiring nondisclosure agreements, should be discussed with the editor no later than at the manuscript revision stage and spelled out explicitly in the acknowledgments.

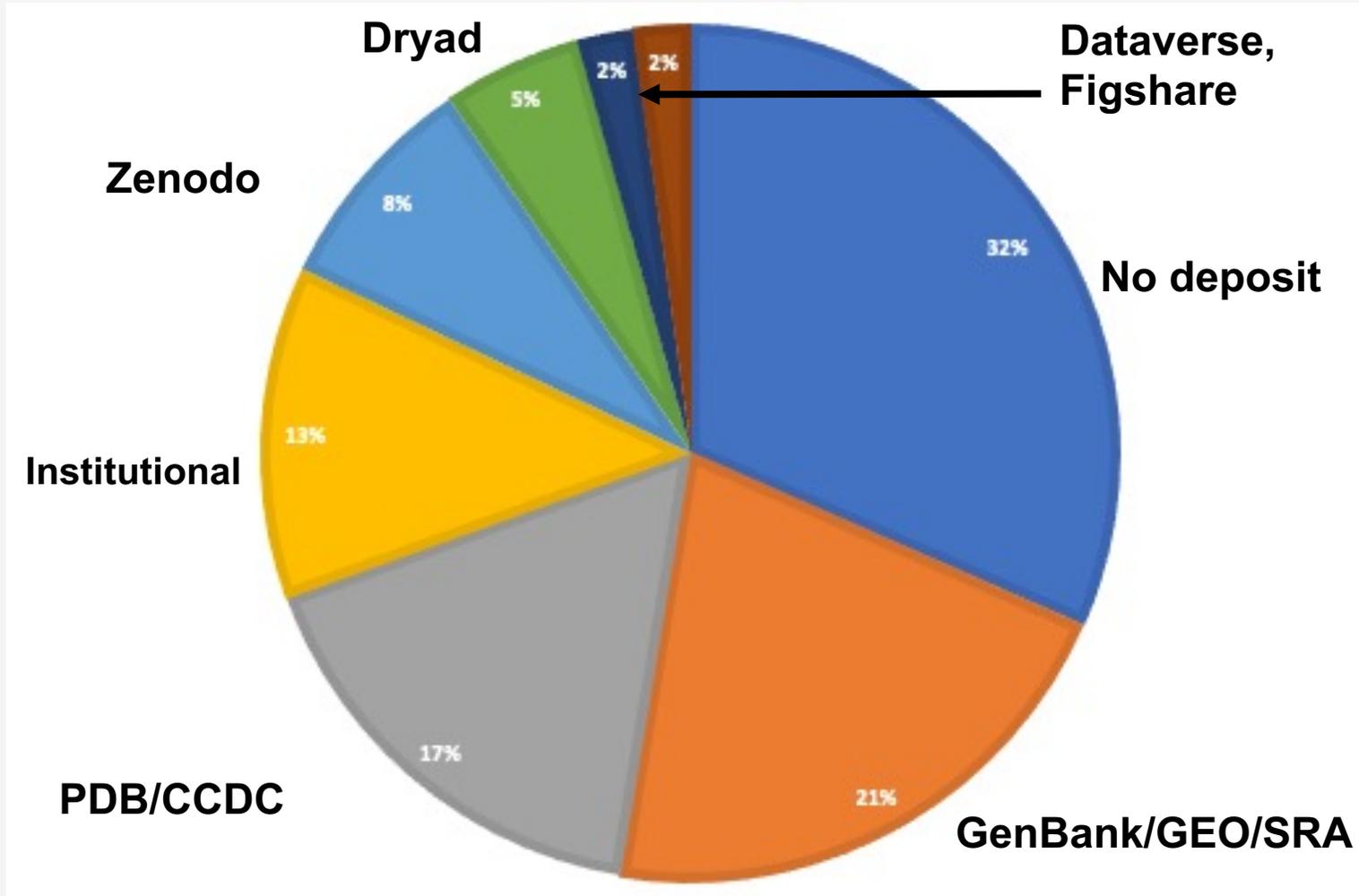
2. Data standards

All data used in the analysis must be available to any researcher for purposes of reproducing or extending the analysis. Data must be available in the paper or deposited in a community special-purpose repository or a general-purpose repository such as Dryad (see **Data and Code Deposition**). Exceptional circumstances requiring special treatment, such as protection of personal privacy or purchase of datasets from third-party sources, should be discussed with the editor as early as possible (no later than at the manuscript revision stage) and spelled out explicitly in the acknowledgments. Problems in obtaining access to published data are taken seriously by the *Science* Journals and can be reported at science_data@aaas.org.

Step 3: Mandating Data Availability Statements

- Applied to papers submitted after January 1, 2018 (which in practice meant published in March or later)
- Most papers stated that data were available in the main text or supplementary materials (which of course depends on how “data” are defined)
- Our biggest remaining challenge is specifying WHICH TYPE OF DATA must be archived for each field of study
- Our second biggest challenge is dealing with third-party constraints that make readers recontact each source in a collated dataset

Step 4: Taking Stock of 2020 papers



12% of papers (93) also included a citation to code archived at Zenodo (an additional 30 GitHub url's also snuck through)

$\frac{3}{4}$ of Institutional data deposits used url's without DOI's (needs to be improved)

1% or less data repositories:
OSF
Materials Data Facility
SEANOE
PANGAEA
National Climatic Data Centers
National Snow & Ice Data Center
International Tree-Ring Data Bank
Earthdata USAP
EarthChem
Neptune Sandbox Berlin

Takeaways

- Crystallographic and genomic repositories have been MANDATED for over a decade, so their prevalence is not surprising
- How can authors be persuaded of the value of field-specific over general repositories?
- Our biggest remaining challenge is specifying WHICH TYPE OF DATA must be archived for each field of study
- Our second biggest challenge is dealing with third-party constraints that make readers recontact each source in a collated dataset

RDA Framework Compliance 1

- Embargoes (explicitly prohibited)
- Data Citation (explicitly mandated, though we do not yet mandate DOIs)
- Researcher/Author Support (offered by individual editors responsible for each discipline, plus a dedicated email address for readers to complain if they cannot access data)
- Data Availability Statements (explicitly mandated)
- Mandatory Data Sharing (explicitly mandated)

RDA Framework Compliance 2

- **Definition of Data (work in progress, especially across different disciplines)**
- **Definition of Exceptions (work in progress, as noted above pertaining to third party constraints)**
- **Data Formats and Standards (work in progress, specified in some fields such as crystallography but not others)**
- **Data Repositories (again work in progress: permanence and accessibility are mandated, several general repositories are recommended, and we link to the AGU/Datacite repository finder, but we don't set criteria such as CoreTrustSeal, which is a problem for Institutional repositories in particular)**
- **Supplementary Materials (we nudge authors toward repositories but still allow data files in the supplement if the authors have a strong preference)**

RDA Framework Compliance 3

- **Data Licensing (we do not yet offer guidance, but we should)**
- **Peer Review of Data (we do this on an as-needed basis, but in the medium term we are aiming to partner with one or more repositories to make the process more routine)**
- **Data Management Plans (we do not yet offer guidance, and are open to discussions about the subject)**

Thank you!

**A public blog post summarizing
our 2020 data and code deposition
and citation record is forthcoming!**

Questions?

@JakeYeston on Twitter

jyeston@aaas.org by email