

COMMENT

SCIENTIFIC RECORD How to credit those who generate data when results are reused **p.30**



SCI-FI New Neal Stephenson novel geeks out on virtual life extension **p.33**

ECOLOGY Paean to the power of algae takes in malaria and blanchmange **p.34**

PUBLISHING Remember when *Nature* poured scorn on preprints? **p.36**

LUCAS JACKSON/REUTERS



Oceanographer David Holland repairs a broken navigation module in Greenland.

Make all scientific data FAIR

Follow the geosciences and demand best practice in publishing and sharing data, urge **Shelley Stall** and colleagues.

Scientific data are burgeoning — thousands of petabytes were collected in 2018 alone. But these data are not being used widely enough to realize their potential. Most researchers come up against obstacles when they try to get their hands on data sets. Only one-fifth of published papers typically post the supporting data in scientific repositories — as has been shown by *PLoS ONE*¹.

Too much valuable, hard-won information is gathering dust on computers, disks and tapes.

Scientists don't share data for many reasons. Those who create data rarely receive credit, and when they do, recognition is often limited to citations (see page 30). Scant support is available for curating data. These issues span all disciplines, but conversations are disconnected.

That's why more than 100 repositories, communities, societies, institutions, infrastructures, individuals and publishers (including the Springer Nature journals *Nature* and *Scientific Data*, see go.nature.com/2wbn4kj) have signed up since last November to the Enabling FAIR Data Project's Commitment Statement in the Earth, Space, and Environmental Sciences for depositing and sharing ►

► data (see go.nature.com/2wv2jxd). The principles state that research data should be 'findable, accessible, interoperable and reusable' (FAIR)². The idea is not new, but aligning this broad community around common data guidelines is a radical step.

In practice, this means that the vast majority of Earth-science journals will no longer accept separate data supplements, which can be hard to exploit. Editors will insist that key data are made available in repositories that support the FAIR principles. These changes in policy and practice elevate data to valuable research contributions rather than files that are shoved in as an afterthought. They promise to open up avenues of scientific discovery and to improve replicability.

The benefits of open and FAIR data are enormous. Sectors of society worth trillions of dollars use geoscience data for operations, products and services³. For example, weather prediction draws on meteorological and other data from around the world. The Global Positioning System depends on real-time observations of solar activity, atmospheric dynamics and Earth's gravity. Earthquake-hazard mitigation and verification of the Comprehensive Nuclear-Test-Ban Treaty use seismic data from instruments worldwide.

We now call on the entire scientific community to implement these practices, and we set out here how the geosciences community achieved them.

BUILDING BLOCKS

The FAIR principles are the culmination of more than 20 years of agreements and actions involving publishers, data repositories, funders of scientific research, researchers and others. The principles recommend that scientific data are: 'findable' by anyone using common search tools; 'accessible' so that the data and metadata can be examined; 'interoperable' so that comparable data can be

analysed and integrated through the use of common vocabulary and formats; and 'reusable' by other researchers or the public as a result of robust metadata, provenance information and clear usage licences.

For example, data sets in the EarthChem Library at the Interdisciplinary Earth Data Alliance are easily found through Google Dataset Search. Data are straightforward to download from a landing page. Data-set formats are aligned with other geochemical, petrological and geochronological data. And they have a long useful life because of their rich metadata on provenance.

Data should be as open access as possible, but sometimes need to be restricted for legal or other reasons: exact locations of observations of endangered species, for example, are restricted, and an approval process must be followed to gain access. Any restriction on access should be spelt out in the data-availability statement of the related paper.

It took just 18 months for the community to adopt, adapt and align with the FAIR data practices. The effort began in 2017 and involved more than 300 stakeholders and six working groups. It was convened by the American Geophysical Union and was supported by the US philanthropic organization Arnold Ventures (previously the Laura and John Arnold Foundation).

This fast pace was possible because many building blocks of data sharing had already been developed. For example, data communities such as the Research Data Alliance and Earth Science Information Partners had vetted and made actionable the data-sharing practices necessary for repositories, researchers and journals. And an alliance of publishers, journals and repositories — the Coalition for Publishing Data in the Earth

"Any restriction on access should be spelt out."

and Space Sciences — had promoted common policies and procedures for publishing and citing geoscience data.

By 2018, all these building blocks had been put together into a unified structure, which the community was keen to implement^{4,5}. The outcomes are formalized as the Enabling FAIR Data Commitment Statement. This contains codes of practice directed at each stakeholder group (repositories, publishers, societies, communities, institutions, funding agencies and organizations, and researchers).

For example, publishers agree to adopt a shared set of author guidelines for storing and citing data. Journals are phasing out the listing of data in supplementary information and are guiding authors to deposit data in repositories that support the FAIR principles. Repositories provide persistent identifiers, curation expertise, landing pages and support for the citation of data in papers. They offer consistent and clear information that is simple to find and access, in formats that are easy to read by humans and machines, with connections to related publications.

CHANGING THE CULTURE

Three big changes are crucial to shift research culture across all disciplines:

Make depositing open and FAIR data a priority for all. Universities, funders, repositories, publishers and societies worldwide need to cooperate to harmonize data-sharing approaches and tools. All journals should demand that data sources are cited and made accessible as an essential part of the integrity of published research. Funders should support leading practices in data management, including long-term archiving of data, especially from publicly supported research. Repositories should track scholarly output to link assertions and evidence.

Stronger mandates and guidance are needed to align these actions. Initial efforts are under way. The report⁶ of the European Commission's Expert Group on FAIR data has set out the steps needed. The Australian Research Data Commons group has produced online training guides and self-help web pages to assist researchers with citing data, samples and software properly (see go.nature.com/2wtuwe8). US research agencies have issued guidelines on how to increase access to scientific data to address the requirements of a 2013 memo by the Office of Science and Technology Policy⁷. Although these initial efforts are encouraging, ongoing recognition and coordination at an international level are needed to align the stakeholders and ensure a common expectation and priority.

Recognize and incentivize FAIR data practices. These should be codified in institutions' reward and tenure processes. The current measurement of a publication's value is



A member of the Hawaii National Guard measures sulfur dioxide levels at a lava flow near Pahoa.

TERRAY SYLVESTER/REUTERS



MARC STEINMETZ/VISUM/EVINE

A researcher accesses sediment cores collected during an ocean-drilling programme.

heavily skewed towards journal citations and poorly reflects the overall value of the research conducted by the authors. Yet the data often have much greater scientific impact than the article to which they pertain⁸.

Researchers should receive credit and recognition for the intellectual effort involved in providing well-documented, useful and preserved data — that is, for practising good science⁹. Societies and academies should include open sharing and FAIR treatment of data explicitly in criteria for honours and awards for exemplary scholars.

Fund global infrastructure to support FAIR data and tools.

The total cost of providing data that meet all the criteria is unknown. Initial estimates are steep, but still small compared with the potential benefits. It will depend on the amount of scientific data involved, and the effort required to comply with FAIR guidelines. The full costs of international FAIR data infrastructures have also yet to be determined. Those parts that can be accounted for do not have stable support. For example, most repositories struggle with the problem of their sustainability.

Researchers should not be expected to bear the entire cost of moving to FAIR data. International coordination of funding is needed. Technical solutions must endure through political and technological transitions, and they must go beyond national boundaries to ensure equal access for researchers from

low- and middle-income countries. Research teams need access to data experts.

Changing cultures takes time and persistence, but the problem is urgent. Progress in the geosciences is encouraging. Some technical problems still need to be solved, but the greatest challenges are organizational and institutional. Other scientists should begin to address them now.

We invite researchers and organizations across all scientific fields to sign up to the Enabling FAIR Data Commitment Statement, assess the current situation in their disciplines, move things forward and report on progress. ■

Shelley Stall is senior director of data leadership at the American Geophysical Union, Washington DC, USA. **Lynn Yarmey** is director of the Research Data Alliance/US Community Development at the Research Data Alliance, Rensselaer Polytechnic Institute, Troy, New York, USA. **Joel Cutcher-Gershenfeld** is a professor at the Heller School for Social Policy and Management, Brandeis University, Waltham, Massachusetts, USA. **Brooks Hanson** is executive vice-president for science at the American Geophysical Union, Washington DC, USA. **Kerstin Lehnert** is Doherty senior research scientist at the Lamont–Doherty Earth Observatory of Columbia University, Palisades, New York, USA. **Brian Nosek** is executive director of

the Center for Open Science, Charlottesville, Virginia, USA, and a professor in the Department of Psychology, University of Virginia, Charlottesville, USA. **Mark Parsons** is a senior research scientist at the Tetherless World Constellation, Rensselaer Polytechnic Institute, Troy, New York, USA. **Erin Robinson** is executive director of Earth Science Information Partners, Boulder, Colorado, USA. **Lesley Wyborn** is adjunct fellow at the National Computational Infrastructure Australia, Australian National University, Acton, Australia. e-mail: sstall@agu.org

1. Federer, L. M. *et al.* *PLoS ONE* **13**, e0194768 (2018).
2. Wilkinson, M. D. *et al.* *Sci. Data* **3**, 160018 (2016).
3. Hanson, B. *et al.* *Eos* <https://doi.org/10.1029/2018E0071991> (2017).
4. Stall, S. *et al.* *Eos* <https://doi.org/10.1029/2017E0088425> (2017).
5. Stall, S. *et al.* *Eos* <https://doi.org/10.1029/2018E0109301> (2018).
6. European Commission. *Turning FAIR into Reality* (European Union, 2018).
7. Holdren, J. P. Memorandum for the Heads of Executive Departments and Agencies: Increasing Access to the Results of Federally Funded Scientific Research (Office of Science and Technology Policy, 2013).
8. Belter, C. W. *PLoS ONE* **9**, e92590 (2014).
9. National Academies of Sciences, Engineering, and Medicine. *Open Science by Design: Realizing a Vision for 21st Century Research* (National Academies Press, 2018).

L.Y., J.C.-G., K.L., B.N. and E.R. declare competing financial interests; see go.nature.com/2xcp4f4 for details.

CORRECTION

The Comment 'Make scientific data FAIR' (*Nature* **570**, 27–29; 2019) wrongly stated that Springer Nature signed up to the Enabling FAIR Data Project's Commitment Statement; only *Nature* and *Scientific Data* have so far done so.