10 June 2019
(Oceanographer David Holland repairs a broken navigation module in Greenland. Credit: Lucas Jackson/Reuters)

Source: Nature - International Jornal for Science (Link at the bottom)

All disciplines should follow the geosciences and demand best practice for publishing and sharing data, argue 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. 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*) 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.

© Copyright 2020 – FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

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 services[3]. 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.

© Copyright 2020 – FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

(A member of the Hawaii National Guard measures sulfur dioxide levels at a lava flow near Pahoa.Credit: Terray Sylvester/Reuters)

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 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

© Copyright 2020 – FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

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 report[6] 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 Policy[7]. 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.

© Copyright 2020 – FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

(A researcher accesses sediment cores collected during an ocean-drilling programme.Credit: Marc Steinmetz/VISUM/eyevine)

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 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 pertain[8].

Researchers should receive credit and recognition for the intellectual effort involved in providing well-documented, useful and preserved data — that is, for practising good science[9]. 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.

--

Source: Nature - International Jornal for Science - https://www.nature.com/articles/d41586-019-01720-7

© Copyright 2020 – FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.