

THIS WEEK

EDITORIALS

SUSTAINABILITY The rush to build dams in the Amazon may be misguided **p.328**

CLIMATE A possible happy ending for US hokey-cokey on Paris Agreement **p.329**



SEA HEAR Fossils show that marine life changed whale ear structure **p.331**

Not-so-open data

Too many funding agencies resist supporting the sharing of data and too many research communities struggle with its practicalities. The result is empty rhetoric and slow science.

Everyone agrees that there are good reasons for having open data. It speeds research, allowing others to build promptly on results. It improves replicability. It enables scientists to test whether claims in a paper truly reflect the whole data set. It helps them to find incorrect data. And it improves the attribution of credit to the data's originators. But who will pay? And who will host?

Only rarely does a research-funding agency step up to both of these plates. Examples include NASA and the US National Institutes of Health, and the European Bioinformatics Institute. The National Natural Science Foundation of China has ambitions to host the outputs of those that it supports. The European Commission hopes to offer such platforms with its European Open Science Cloud. The UK Data Archive for the social sciences and humanities, and DANS, the Netherlands Institute for Permanent Access to Digital Research Resources, represent other good models of support from governments.

But in too many cases, government agencies lack the funds to build platforms for data sharing and resist taking responsibility for such infrastructure. They may hope that universities will host data, but the development of institutional repositories is patchy, and to rely on them is effectively to discourage common data standards and curation.

There are commercial data platforms, including figshare (which shares a common owner with *Nature*, through Holtzbrinck Publishing Group). Given their usefulness, it is surely misguided for funding agencies — for instance, the Swiss National Science Foundation — to prohibit their use by grant-holders. There are also not-for-profit repositories such as Dryad.

As *Nature* well knows, being a host — or publisher — of data is expensive. Keeping a platform technologically up to date is costly, as are data validation and curation. The running costs of the preprint server arXiv in 2017 are about US\$1.3 million, for example, and the 2015 budget of the UK Data Archive was about £5.5 million (\$8.2 million). For too long, public discussions have overlooked the true costs of data openness. More tangible support from governments and funders would work wonders.

Bottom-up motivation from researchers to share data is also crucial — and needs encouragement.

Genomics and structural biology have an honourable history of insisting on the prompt deposition of open data and providing facilities for doing so. Other communities also have strong customs surrounding data ownership. For example, astronomers who have developed instruments for satellites often have proprietary access to new data for a year, but many astronomical facilities create their own rules. Even when journals insist on immediate access for readers to the data included in a research publication, the full data set and the software required to analyse it may be kept from readers for months. Given the diversity of data and conventions, it is up to funders, researchers and journals to keep up the pressure towards the openness of complete data sets and any source code required to use them.

So which fields need to raise their data-access game? *Nature* suggests that the geodesy and seismology communities should consider reducing their current two-year embargoes. The microbiome community places great value on open data but, as a relatively young field, is struggling to establish standards.

Thumbs up for two communities that are making progress in this realm. In pathogen genomics, the authors of the Zika virus genome papers we publish in this issue (see pages 401, 406 and 411) made the sequences openly available as soon as they were generated.

“For too long, public discussions have overlooked the true costs of data openness.”

Credit should also be given to palaeontologists in their pursuit of an open strategy for 3D data. A recent paper, ‘Open data and digital morphology’ (T. G. Davies *et al. Proc. R. Soc. B* **284**, 20170194; 2017), proposes

best practice for the creation, storage, publication and dissemination of large 3D data sets, including recommended file formats and data repositories. The conclusion is that 3D data should be available at the time of article publication, accompanied by as much detail as possible on its nature and the circumstances of its collection.

Where researchers establish clear standards and repositories, *Nature* will be delighted to help mandate their use as a condition of publication. Where such coherence is lacking, we necessarily take a more piecemeal approach. Despite years of discussion, funders, researchers and journals have much work to do to improve the transparency and reproducibility of research by means of data accessibility. ■

Speak now

Seize the chance to comment on US regulation of genetically engineered plants and animals.

In January this year, two US agencies proposed the first substantial overhaul in 30 years of how they regulate genetically altered crops and livestock. Some plant scientists expressed relief. Some animal researchers used more colourful language.

The proposals — one to govern plants, the other to govern animals — came to wildly different conclusions. The US Department of Agriculture (USDA) suggests that many plants whose genomes have been altered by a single DNA letter change should not need approval before being released in the field. However, the US Food and Drug Administration (FDA) contends that animals whose genomes have been similarly changed might have to go through a rigorous evaluation before being released onto the market.

For the two agencies to evaluate the same problem and come to