

A statement about data

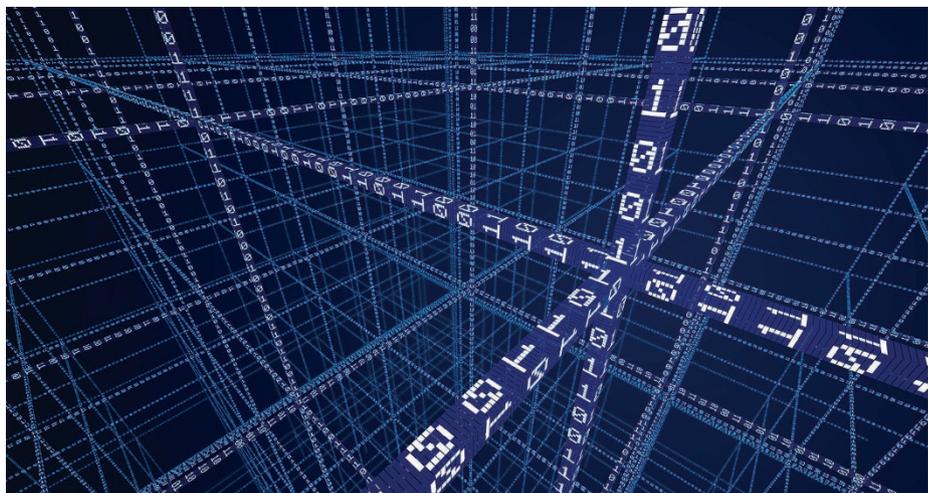
Nature Physics now requires its published papers to include information on whether and how their underlying data are accessible to others.

Over the past few months, eagle-eyed readers of ours may have noticed a new section appearing at the end of our papers: a data availability statement. This is the result of a trial we have conducted alongside four other Nature journals — *Nature Cell Biology*, *Nature Communications*, *Nature Geoscience* and *Nature Neuroscience* — to build on our support for authors wishing to make their data available to other researchers.

The statements report the availability of the data that is necessary in order to interpret and replicate the findings in the paper. At a minimum, we expect authors to confirm explicitly that all the relevant data are available upon request, making clear any restrictions on availability (perhaps due to privacy limitations or third-party control). However, the statements may also provide details about publicly archived data sets and, whenever possible, we now encourage authors to cite any data sets that have a digital object identifier (DOI) assigned to them.

As one might expect, the pilot confirmed different approaches and practices to data sharing in different disciplines, and it also uncovered certain practical obstacles resulting from this. Although most authors are aware of funder and institutional mandates encouraging and, in some cases, even compelling them to share their data, there is often a lack of obvious, public, community repositories for them to use. Raising awareness of existing repositories is therefore one challenge. But increasing the awareness that data deposition can enhance the visibility, reproducibility and reuse of published research, and that data citation can increase the recognition of those who create and share data, is also important.

Scientists in certain areas of physics are, of course, already familiar with data sharing. Astronomers, for example, have long been cataloguing and archiving data related to specific telescopes, space missions and consortia. In general, these adhere to policies that require most of the data to become freely available after some minimal proprietary period — the release of a year's worth of observations of the Milky Way from the European Space Agency's Gaia probe being a recent, spectacular example of this practice



© ANDRZEJ WOJCIK / SCIENCE PHOTO LIBRARY / GETTY IMAGES

(see page 896 for Marios Karouzos's take on the data). It therefore comes as no surprise that, so far, it is our astronomy papers that contain the most detailed data availability statements.

Another constituency of researchers making use of large-scale central facilities — albeit ones focusing on completely different length scales — are condensed-matter physicists. Scattering and spectroscopy techniques based on neutrons, X-rays and electrons continue to provide an unrivalled degree of insight into the fundamental properties of materials, but they often require access to a neutron source or synchrotron radiation in order to be performed. These facilities tend to be hosted by national or international laboratories and access to them is increasingly conditional on making the data from the experiments they allow available to others. Most North American and European facilities even provide a data repository of their own and, though it is comparatively early days for them, we hope data availability statements will raise their profile.

Of course, many of our authors work in the scientific 'cottage industry' of university or company laboratories and they have access to far more limited resources. How can these scientists share their data? Some may find that their own institutions encourage and support data archiving, but repositories such as figshare (www.figshare.com) and Dryad (www.datadryad.org) also make it possible

to make data outputs available in a citable, shareable and discoverable manner.

Be that as it may, it is fair to say that the practice of data sharing is still a work in progress. While certain fields are leading the way in terms of best practice, for many others there is still a gap between the expectation and the reality of seamlessly sharing data alongside publications. Publishers can and should do more in this regard. The publication of data descriptors (for example in the Nature Research journal *Scientific Data*) helps to provide detailed descriptions of experimental, observational, computational or curated data, but better tools to integrate papers with existing repositories are sorely needed. Nevertheless, encouraging data sharing in a way that reflects the circumstances of respective specialist communities is surely a step in the right direction.

So has the trial been a success? We believe it has. Earlier last month, *Nature* took the next natural step of this initiative, which is to announce our new policy on data availability statements and data citations (*Nature* 573, 138; 2016). Full details on the policy, which will eventually cover all Nature-branded journals, are available at <http://go.nature.com/2bf4vqn>. We expect that offering consistent information on data availability in our papers will promote data reuse by future researchers, and we are convinced it represents a small but decisive step in the right direction for transparency in research. □