

It's all about data

Research data comes in various forms and levels of significance. Finding the best way to share all of the results of a research project can be difficult, but new ways are constantly emerging.

Scientists disseminate their work by writing and publishing scientific papers, but this finished product can conceal a wealth of effort and information. Behind the text and figures is the data itself, which has been recorded, analysed, interpreted and, eventually, summarized in graphs and images. The process is necessary, but it can mean that accessing the underlying data is not straightforward.

A condition of publishing in *Nature Nanotechnology*, and in the other *Nature* journals, is that authors are required to make data and associated protocols promptly available to readers on request (<http://www.nature.com/authors/policies/availability.html>). A few months ago, some of our sister journals in the life sciences also introduced the possibility of directly linking graphs in a paper to the source data, which are included in the Supplementary Information (P. M. Gubser, *Nature Immunol.* **14**, 1064–1072; 2013). We are hoping to implement this functionality at *Nature Nanotechnology* soon. In the meantime, we encourage our authors to include the underlying data for figures in a paper as part of the supplementary information, as has been done on a number of occasions before (V. S. Pribiag *et al.* *Nature Nanotech.* **8**, 170–174; 2013).

Scientists should strive to provide a full account of their research results, including those that do not necessarily fit the main message of the paper.

The data reported in a paper is often, however, only a fraction of what was obtained during an experiment or calculation. And despite the time and resources required to obtain this information, the majority of the data can remain unused or forgotten, and inaccessible to the wider scientific community. This has led to calls from scientists and funding agencies to make all types of data more openly available and the emergence of a variety of public repositories, where the information can be stored and accessed by others.

An important example of data that often remains hidden from the community is negative data. In a Commentary in this issue on page 693, Leonie Mueck examines the topic and analyses how often negative results are reported across various disciplines and the reasons for not reporting them. It is a complex issue, especially as there are several types of data that can be defined as being negative and the definition depends strongly on the discipline. In some cases, part of the data can be considered as null, in the sense that they do not produce any expected effect. In some cases, they may not produce any pattern that can be explained by a simple model. Either way, the results are never published and the data forgotten. As Mueck argues, scientists should instead strive to provide a full account of their research results, including those that do not necessarily fit the main message of the paper.

One of the problems of reporting negative data is the scarcity of venues that focus on these types of result. There are, however, exceptions such as the *Journal of Negative results in Biomedicine* or the *Journal of Negative Results — Ecology and Evolutionary Biology*. A slightly different approach is taken by the *Journal of Unsolved Questions*, which apart from offering the possibility of publishing negative results, discusses open questions in science and reflects on current publication practices. The online repository Figshare (which, like *Nature Nanotechnology*, is part of Macmillan Science and Education) also offers the possibility of publishing negative results in a way that can be shared and cited.

Apart from negative results, another difficulty in data publishing concerns making large sets of data accessible. In such cases, the data have to be presented in a coherent way so that they can be easily understood. Furthermore, the researchers involved in acquiring the data and then organizing them in an accessible way should obtain recognition for their efforts.

These are some of the reasons that Nature Publishing Group is launching the online-only publication *Scientific Data*. This open-access publication will go live in Spring 2014, and is now starting to



© PHOTOGEN/ALAMY

accept submissions. Rather than publishing scientific papers, *Scientific Data* will publish descriptions of datasets, including the methods and protocols used and details of the community-recognized repositories in which all the original data can be found (sample publications can be found at <http://www.nature.com/scientificdata/>). These 'data descriptors' are peer reviewed and are citable with a unique digital object identifier, and will allow authors to receive credit for depositing and sharing data. The data descriptors can be related to results published in a scientific paper, or can be standalone publications. Submissions to *Scientific Data* will be evaluated based on the technical quality and reuse value of the datasets, not specific interpretations. Authors will have a place and an incentive to publish some of their valuable 'hidden' datasets, including some that might be considered 'negative' in a traditional research journal.

Initially, *Scientific Data* is focusing on life, biomedical and environmental sciences, but they will be progressively expanding into other scientific disciplines, including nanotechnology. Regardless, the continuing push to make data more open and accessible should benefit all of science. □

Corrected after print: 20 November 2013

Correction

In the version of the Editorial 'It's all about data' originally published (*Nature Nanotech.* **8**, 691; 2013), '*Journal of Questions*' should have read '*Journal of Unsolved Questions*'. Corrected in the PDF and HTML versions after print 20 November 2013.