

# Data take centre stage

We are updating our editorial policies to further encourage authors to make their data publicly accessible. Publishing Extended Data figures and source data online will also ensure that data are given a more prominent role.

A Data availability statement for accepted articles has been requested by the Nature Research journals for the past three years<sup>1</sup>. In this statement, our authors declare how the data behind their published research can be accessed by interested readers, and disclose any potential restriction limiting data sharing. This initiative, aiming at increasing the reproducibility of our papers, has been seamlessly accepted by researchers, who have promptly added this declaration to their manuscripts submitted to the Nature titles. However, in the vast majority of cases data have remained locked up in the authors' drawers, and allowed to see the light only 'upon request'.

In their contribution to the Why it Matters column, Natasha Noy, computer scientist at Google Research, and Aleksandr Noy, materials scientist at the Lawrence Livermore National Laboratory in California, USA, reveal that a disappointingly small fraction of the papers published in the first 10 issues of *Nature Materials* in 2019 provided publicly accessible links to data repositories. Yet for materials scientists there are plenty of good reasons to share data — for instance, their results can be merged with those of other labs to create larger datasets, and used for further statistical analysis or for a better training of machine learning algorithms for materials discovery. In addition, several research funders already mandate the public availability of data generated by the researchers they financially support; similarly, publicly funded large-scale facilities require their users to give unrestricted access to their results. Above all, sharing promotes transparency and pushes scientists to scrutinize their own laboratory practices related to the production, presentation and storage of research output.

Public repositories, which provide digital identifiers to facilitate proper credit attribution to authors and ensure permanent preservation of the data, are widespread on the Internet. For instance, a list of repositories endorsed by the relevant scientific communities for specific datasets can be found in the Nature Research policies on availability of data (<https://www.nature.com/nature-research/editorial-policies/>



Credit: Peter Cade/Getty

[reporting-standards#availability-of-data](#)) — and the list is rapidly growing. Discovery of these datasets is also becoming increasingly easier thanks to the development of dedicated search engines, such as Google Dataset Search (<https://toolbox.google.com/datasetsearch>), developed by Natasha Noy and her colleagues.

To make the whole process work efficiently, an extra step is needed from researchers. They have to prepare their data in a shareable form and add metadata that describe the data measured, the techniques and experimental conditions used, and other information key to understanding the content provided. Yet Natasha and Aleksandr Noy explain that several repositories are streamlining this procedure by generating such metadata automatically from the details provided by the authors during upload, and other tools are available to generate them independently.

Now, to further encourage our authors to make that extra step, our journals have modified the requirements for the preparation of the Data availability statement on acceptance of scientific articles for publication. If data, as well as custom computer code used in the article, can be made available only upon request, we will ask for an explanation of the reasons in the published statement. We hope that reflecting on any assumed barriers to sharing will help authors remove them whenever possible, and transparently motivate those cases where specific roadblocks exist that cannot

be overcome (for instance, for the presence of sensitive information or personal data that cannot be anonymized).

We are also providing more space for data in the web pages of *Nature Materials*. Following the example of *Nature*, we will now allow the publication of up to ten Extended Data figures, which will be integrated in the HTML version of the paper and added to the main PDF online (but not included in the printed version). This will give these figures more visibility with respect to the results presented in the Supplementary Information, and as such researchers are welcome to use them to report key supporting evidence that strengthens the claims of the paper. Importantly, authors are also encouraged to provide 'source data' for all the figures reported in the main manuscript and Extended Data. These could be unprocessed source images for electrophoretic gels and blots (which we already asked to include in the Supplementary Information, <https://www.nature.com/nature-research/editorial-policies/image-integrity>) or microscopy data, as well as tabulated numerical data underlying the graphs presented in the article. A link to the related source data, which are published online as individual supplementary files, will be provided at the bottom of each figure in the HTML version of the paper. Hopefully this will make life easier for researchers that wish to perform additional analyses on published data and that, until now, had to use dedicated software to extract data points from our charts. Both Extended Data figures and source data will not be mandated — authors will decide whether to make use of these features or not. Like with the use of public repositories, however, we expect a large uptake from the whole materials science community.

Data are the backbone of every scientific discovery; they are the ground on which researchers confront opinions, debate theories and build collaborations. Unlock them from your drawers and hard drives, and give them the visibility they deserve. □

Published online: 18 December 2019  
<https://doi.org/10.1038/s41563-019-0574-2>