

The importance and challenges of data sharing

Nature Nanotechnology will work with authors and reviewers to establish the best practises to make data available and usable by a wide scientific community.

Data are the foundation of scientific progress. They are obtained with considerable efforts by researchers, mostly through projects supported by public funding. However, their purpose is often limited to the production of scientific publications and, unfortunately, the vast majority of data are not utilized and instead remain stored in some electronic storage device.

Examining existing data is beneficial for a number of reasons. Scientists building their research on previous work can optimize acquisition protocols and design experiments that complement previously obtained datasets. Access to data — and understanding the way in which they were acquired — is essential for theoreticians attempting to provide a mechanistic explanation of observed phenomena. In fields such as environmental nanotechnology or nanomedicine, the analysis of wide sets of data can help to create a big picture of the effects of nanomaterials, which is necessary for product development and for the evolution of regulations. Finally, data availability is crucial for research transparency.

Like all Nature journals, *Nature Nanotechnology* requests, as a condition of publication, that all data relevant to the conclusions of a paper are made available by the corresponding author upon request. Authors are also encouraged to submit files for the data points included in every figure, as source data. But, to increase transparency and usability of data, we also strongly encourage the deposition of all the data necessary to support the conclusions of papers in public repositories, together with a description of how the data have been obtained.

In March 2017, all Nature journals introduced a data availability statement to be published with every paper (<https://www.springernature.com/fr/authors/research-data-policy/data-availability-statements>), in which the authors must specify whether the data related to the paper can be accessed publicly — and if so, a reference provided — or if they can only be obtained from the corresponding author upon reasonable request. Of all the papers published with a data availability statement up until February



Credit: Westend61 GmbH/Alamy Stock Photo

2020, 88% provide a statement that declares the data to be available from the authors upon request. The remainder of statements are quite varied. In some cases, only part of the data is made publicly available in repositories. In others, the data is included in the supplementary information. Only in a small handful of cases are all data deposited in a public repository.

At this stage, we do not know enough to determine the reasons behind such a small number of cases in which the data is deposited in public repositories. In some fields it may simply be a matter of establishing common practices. But we are aware that sharing data also poses challenges as it requires the authors to organize the raw data and relate it to the data presented in the figures in a way that is understandable to (and ideally reusable for) others. Given the complexity and substantial amount of data in *Nature Nanotechnology* papers, this can be a very time consuming and daunting task, and demands a high level of data organization from the beginning of the project and the experiment. We also appreciate that there simply isn't a universal definition of what constitutes the data that support the conclusions of a paper. For example, raw data could be useful in some cases, but difficult to interpret in others. This can vary from one paper to another in the same field, let alone for studies in different fields of nanotechnology.

We are convinced that making data generated during a scientific study available and easy to interpret and reuse is essential for the future of science. For specific areas, like protein or DNA sequencing, structured and commonly accepted databases already exist. In the absence of repositories specifically designed for their area of research, authors are recommended to use general repositories, such as figshare, Zenodo and Dryad. They can also contact the research data helpdesk for independent advice (<https://www.springernature.com/de/authors/research-data/helpdesk>).

We will continue to encourage our authors to engage with public deposition and with data description. In line with a step taken by other *Nature* Journals, we are now asking authors of accepted papers for which data are not publicly available to explain why that is the case. We hope that this will help to establish whether the reason is simply a choice or if it is due to concrete obstacles. In addition, starting from February 2020 editors at *Nature Nanotechnology* are asking reviewers to provide advice to authors on which data would be useful to share. We plan to review the information gathered before deciding whether any specific action should be taken in terms of publishing policies. □

Published online: 10 February 2020
<https://doi.org/10.1038/s41565-020-0646-0>