Dealing with data

Nature Materials now requests that all original research articles contain a Data Availability Statement declaring the accessibility of the data and where it can be found.

Continuous development is an essential objective of a publisher committed to support the progress of research and the dissemination of scientific discoveries. Several projects are pursued by Springer Nature in order to meet the needs of authors and research communities — from cross-journal initiatives like SharedIt, the platform that enables authors and subscribers to publicly share free-to-read links to research papers published by journals in the Springer Nature portfolio[1], to pilots such as the trial of *BMC Psychology* to peer-review manuscripts that do not disclose the study results, as an investigation of whether such practices would reduce publication bias[2].

One of the established conditions for publication in Nature journals, consistent with the general policies of the publisher Springer Nature, is that datasets enabling the reproduction or reuse of articles' findings are made accessible to the readers[3]. This is why since the end of October 2016, and following the favourable outcome of a probation period involving a subset of the Nature titles, *Nature Materials* has required that all published articles reporting original research contain a Data Availability Statement, in which the readers are clearly informed where the underlying data can be found[4]. An analogous declaration has already been in place for a while for works that use custom computer programs — in these cases it is expected that authors clarify where the code of the software can be found.

Ideally, and in particular for large datasets, data should be submitted to type-specific public repositories and their accession numbers or, if applicable, digital object identifiers (DOIs) should be declared. For an interdisciplinary subject such as materials science, there is naturally a variation in the practices and relevant online archives for data deposition across the different communities. Deposition of certain types of data is mandatory; for instance, macromolecular structures can be collected in the Worldwide Protein Data Bank[5], and small-molecule crystallographic data can be stored in the Cambridge Structural Database[6]. Other databases that may be relevant for branches of materials science are the NoMaD repository[7], which contains calculated material properties, caNanoLab

©MICHAEL ZWAHLEN/EYEEM/GETTY

for descriptions of nanomaterials examined for biomedical applications[8], and PubChem[9] (linking three databases) for characterizations of chemical substances. Even experimental protocols can be freely shared at the Protocol Exchange[10].

If no community-recognized data management option is available, general resources such as Figshare[11] and Dryad Digital Repository[12] are recommended. These are also integrated with the submission system of *Scientific Data*, the open-access journal that publishes descriptions of research datasets, and provides a further list of assessed, subject-suitable data archives[13]. Under certain circumstances, the nature of the data is such that it cannot be freely released — in such cases, a confirmation should be given that data are available from the authors upon reasonable request.

In view of recent examples of the significance of accessible data in periods of disease outbreaks leading to public health emergencies[14], or flags raised over the unexpected degree to which phylogenetic analyses cannot be reproduced[15], it is difficult not to agree about the importance of clear and precise reporting of materials and experimental methodology, as well as the value of facile access to structured raw data. Due to the more immediate impact on human health or humans being the source of investigated materials (as in the case with the Human Genome Project), driving forces behind open data initiatives have often been found in the life science communities. Also, research groups utilizing internationally supported resources (for example, particle physics accelerators or space telescopes) have traditionally been sharing their data. Moreover, the demand from funding organizations that data produced in projects benefitting from their financial support are made publicly available is increasingly common.

In materials science, where modelling is recognized to provide vital momentum for translating basic research into industrial applications, there is awareness that theoreticians today are limited by their access to experimental data[16]. Likewise, one of the strategic activities of the Materials Genome Initiative — launched to expedite the discovery and manufacturing of advanced materials — is to support a practicable infrastructure for materials data[17]. Good examples of the productive outcome of organized data sharing are resources such as the Materials Project, AFLOW and the Open Quantum Materials Database, where properties of hitherto non-synthesized compounds are predicted based on structural information of known materials.

With this request from a publisher's side for an explicit specification of where the data underlying scientific discoveries can be found, authors will hopefully be encouraged to more thoughtfully and methodically document and secure a wide data access, for the benefit of transparent and reproducible research practices that facilitate data reuse and collective scientific advancement. ❐

References

1. Springer Nature's commitment to content sharing. *Springer Nature* http://go.nature.com/2g2KYp6 (2016).
2. New 'results free' peer-review process to be piloted for academic publishing. *Springer Nature* http://go.nature.com/2gULBD3 (2016).
3. Availability of data, material and methods. *Springer Nature* http://go.nature.com/2gIci08 (2016).
4. Data availability statements and data citations policy: guidance for authors. *Springer Nature* http://go.nature.com/2bf4vqn (2016).
5. http://www.wwpdb.org/
6. https://www.ccdc.cam.ac.uk/
7. http://nomad-repository.eu/cms/
8. https://cananolab.nci.nih.gov/
9. https://pubchem.ncbi.nlm.nih.gov/about.html
10. http://www.nature.com/protocolexchange/
11. https://figshare.com/
12. http://datadryad.org/
13. http://www.nature.com/sdata/policies/repositories
14. Movement for open data during emergencies intensifies as Zika spreads. *World Health Organization* http://go.nature.com/2gYFSfR (2016).
15. Drew, B. T. *Nature* **493,** 305 (2013).
16. *Nat. Mater.* **15,** 365 (2016).
17. *Nat. Mater.* **12,** 173 (2013).