

Where are the data?

Here, we announce two policy changes across Nature journals: data-availability statements in all published papers and official Worldwide Protein Data Bank (wwPDB) validation reports for peer review.

As the research community embraces data sharing, academic journals can do their part to help. Starting this month, all research papers accepted for publication in *Nature* and an initial 12 other Nature titles, including *Nature Medicine*, will be required to include information on whether and how others can access the underlying data.

These statements will report the availability of the ‘minimal data set’ necessary to interpret, replicate and build on the findings reported in the paper. When applicable, they will include details about publicly archived data sets that have been analyzed or generated during the study. When restrictions on access are in place—for example, in the case of privacy limitations or third-party control—authors will be expected to make this clear.

The new policy (full details of which are available at <http://go.nature.com/2bf4vqn>) builds on our long-standing support for data availability as a condition of publication. It also extends our support for data citation, the practice of citing data sets in reference lists in a similar way to citing papers. Authors are encouraged to cite data sets that have digital object identifiers (DOIs) assigned to them.

The introduction of data-availability statements follows a trial at five Nature journals—*Nature Cell Biology*, *Nature Communications*, *Nature Geoscience*, *Nature Neuroscience* and *Nature Physics*—that began in March 2016. The pilot confirmed differences in the culture of data sharing and access among different disciplines, and that the lack of obvious public community repositories can pose a substantial barrier to public data deposition. Nevertheless, even in disciplines that are not yet able to embrace openness and sharing, there is increasing awareness and appreciation that data deposition can enhance the visibility and reuse of published research, and that data citation can increase the recognition of those who create and share data.

This new policy will be implemented across the diverse range of Nature journals by early 2017. We expect that its implementation will shed more light on the reasons for disciplinary differences in data sharing, identify challenges and help to promote the practice more widely. It’s not just journals. A broad drive across the research, funding and publishing communities is under way to make the availability of research data more transparent. Funders, for example, are also introducing data-availability statements. The seven UK research councils require their grant holders to include such statements. And the US National Institutes of Health is asking researchers to provide management plans for their research data.

We expect that offering consistent information on data availability in our papers will promote data reuse by future researchers. And when public data archiving is a mandatory requirement of journals, there is some evidence that including data-availability statements with persistent

links to data in published articles is an effective approach to ensuring public data availability and policy compliance (T.H. Vines *et al.*, *FASEB J.* 27, 1304–1308, 2013).

This new policy follows the launch, in July 2016, by our publisher Springer Nature, of an ambitious project to introduce and standardize research data policies across all of its journals (see <http://go.nature.com/2by6l6x>). The project sets out a defined common framework for data policy—which Nature policies align with—that enables different journals to encourage data sharing in a way that reflects the circumstances of respective specialist communities.

Requesting official wwPDB validation reports

Here, we also announce a change regarding the deposition of macromolecular structures in the wwPDB. For over three years, this journal has asked authors of manuscripts describing new atomic structures to provide validation reports from the wwPDB’s validation server, as a requirement for peer review (*Nat. Struct. Mol. Biol.* 20, 533, 2013). These reports give reviewers more information about the quality of the structural models presented. Our authors have readily complied, and reviewers have come to expect the reports as part of the manuscript files.

We are now taking a further step and are requesting official wwPDB validation reports for peer review. These reports are made available by the wwPDB after data deposition (<http://www.wwpdb.org/validation/validation-reports>). Other Nature journals will soon follow suit.

Our authors may ask why we are making this change. After all, the wwPDB continues to improve the validation server, and the reports for crystal structures are now very thorough. We actually believe that the appropriate question is, why not? If the data are not ready for deposition, they are not ready for peer review either. The change does not create an additional burden to researchers, because they already must deposit the data into the wwPDB at some point; we are simply asking that the deposition be done earlier. Depositing data at the beginning of the editorial process allows sufficient time for the entries to undergo annotation and biocuration by the wwPDB, thereby ensuring that they are ready for release by the time of publication. Finally, as of this spring, wwPDB depositors can choose to suppress the entry’s title and authors until public release (see announcement at <http://www.wwpdb.org/news/news?year=2016#5764490799cccf749a90cdf0>). Thus, researchers no longer need to worry about potentially alerting their competition when they deposit their data.

As always, we welcome feedback from the community on the journal and all editorial matters. If you have questions or comments about the policy changes described here, please feel free to contact us at medicine@us.nature.com.