

Source-ful science

Sharing source data—the actual measurements and unprocessed images behind the graphical representations used in figures—helps to ensure transparency and reproducibility of research results. We urge our authors to submit and share the source data with their published papers.

The ‘reproducibility crisis’, a growing realization that the results of many experiments reported in the literature cannot be replicated and that the conclusions based on them may be unfounded, harms the integrity of the scientific process and undermines public confidence in science. How to improve reproducibility is being actively discussed by the research community, funding bodies and journals, and a number of initiatives to increase transparency in methods and data reporting have been rolled out (*Nature Special on Challenges in irreproducible research*, <http://www.nature.com/news/reproducibility-1.17552>).

In research articles, data are often, if not exclusively, presented in the form of figures, in which charts, graphs or representative blots are used as convenient visual depictions of the results obtained. However, figures usually do not provide the actual measurement values or original images.

Yet, access to the data that are at the base of a scientific paper is crucial not only to ensure transparency and reproducibility but also to allow for data reanalysis and reuse. This is already common practice for specific types of large-scale data sets for which standard public repositories exist. For example, Nature Publishing Group journals require deposition of deep-sequencing and microarray data, and the datasets must be released prior to or upon publication. The same holds true for atomic structures, which must be deposited into the Protein Data Bank. All Nature Publishing Group policies on data availability can be found at <http://www.nature.com/authors/policies/availability.html>.

Nature Publishing Group authors have also been encouraged to provide ‘source data’—minimally processed versions of the data used to generate the figures. Indeed, we have followed the initiative of our sister journal *Nature Cell Biology*, and original, nonmanipulated and uncropped images of gels, autoradiographs and blots are now routinely published with *Nature Structural & Molecular Biology* papers, as supplementary data files.

However, source data also include the individual numerical values and measurements behind the graphs presented in figures. This feature is seldom used by *NSMB* authors. Some researchers might question whether sharing their ‘small-scale’ research data actually serves any purpose and how it can be useful to themselves, their peers or the scientific community at large. We argue here that source data are valuable for multiple reasons.

First, source data reflect the experimental design, which is crucial for the interpretation of the results but is often reported in an incomplete manner. Textual data descriptions in either figure legends or the methods section of a paper are frequently ambiguous. Often, it remains a mystery as to which exact data points were used, and in what way, to calculate the means and errors depicted in the graphs. Source data provide this information in a comprehensible manner and thus ease the interpretation and facilitate the reproduction of published results.

Second, source data provide an efficient means to publicly archive data linked to the research paper. Upon request, authors may be required to fully disclose and provide all data that have led to the conclusion of a study, even years after publication. This becomes particularly relevant if questions ever arise regarding data integrity, and compliance is necessary to avoid lingering concerns that could lead to a retraction of the publication. With the turnover inherent in academic research labs, it can be difficult for principal investigators to keep track of all data over long periods of time, and submitting source data allows efficient and transparent public archiving.

Last, but not least, source data enable reanalysis and potential reuse of the data by the community and thus lead to a wider dissemination of the work. The raw numbers are essential for anyone seeking to build on the results and to integrate them in a quantitative manner. For example, any study modeling a biological process requires experimental input values, which are often drawn from published results. In addition, combining source data points from independent studies might lead to more precise calculations of parameters or add statistical power. New ways to integrate datasets are continuously being developed, and source data could have potential applications that might not be anticipated by the authors.

We appreciate that putting the data into an easily assessable format does require some effort on the part of the authors, who may feel that this is yet another hurdle to overcome in the publication process. We make the case that the additional work is worthwhile, in light of the benefits accrued, and is not as onerous as it may seem. After all, when preparing the figures for a manuscript, researchers have the data at their fingertips. Source data can be easily submitted to *NSMB* in tabular form (as spreadsheets in .xls, .xlsx or .csv formats) and uploaded alongside the manuscript, just like any other supplementary material file. No extensive reformatting from the original data should be required, although labels should of course be kept consistent between the source data files and the actual figures. Ideally, source data should be provided at the submission or revision stage, so that they are part of the peer-review process, but they can also be added after the manuscript is accepted in principle. Ultimately, the data will be made available to readers from links in the figure legends.

In an era in which reproducibility is a major challenge to the progress of science, we strongly encourage our authors to increase data transparency and to share the source data in their research articles. Any comments and suggestions regarding how *NSMB* can improve the submission and presentation of source data in our papers, to benefit authors and readers alike, will be welcome at nsmb@us.nature.com. ■