Correspondence on data

As a result of feedback from the research community, we are strengthening our encouragement for authors to share a certain amount of data with their papers.

In February, we published an editorial (*Nat. Phys.* **15**, 107; 2019) outlining our thoughts on the issues of data sharing and code availability, and asked for feedback from the community. This month, we are publishing some of the responses we received as a representative sample of the views that we have encountered. As a result of this feedback and numerous other conversations, we will start encouraging authors to share a certain minimum level of data when they publish in *Nature Physics*.

On the issue of sharing data, there seems to be a consensus that more openness is a good idea. However, it's clear that the requirements and expectations of different communities and individuals still differ quite widely. Evidence for this is given by Lisa Glaser and co-authors in their [Correspondence](#) describing the *Encyclopedia of Quantum Geometries*. The feedback they received from potential users of the service highlights their disparate storage requirements — from kilobytes to hundreds of gigabytes — but an impressive 88% of their respondents were in favour of sharing their data alongside papers, and the same proportion said that open data would be useful for their own research.

In his [Correspondence](#) Jacopo Bertolotti reinforces the idea that community-specific responses are vital. He points out that sharing the data that go into the figures of the published version of a paper is technologically straightforward and not a big time investment for the authors. After all, they have to have these data to feed into the plotting software. However, the degree of sharing of raw data, more processed data, or large datasets from which averages or correlations are extracted is more complicated and a one-size-fits-all solution is probably not possible.

In the context of large experimental facilities such as the Large Hadron Collider at CERN, Matthew Strassler and Jesse Thaler argue in their [Correspondence](#) that sharing of the large datasets is imperative for science, as otherwise researchers not formally related to the collaboration will have no access to the only data of its kind.

In contrast, the issue of sharing code seems much more complex. In his [Correspondence](#) Gergely Zaránd points out that the loss of competitive advantage that would be felt by mandated code availability

Data servers at CERN. Credit: EThamPhoto / Alamy Stock Photo

would further entrench inequalities of different research groups. This is a view that should be given careful consideration. Philip Moll backs up this stance in his [Correspondence](#) by suggesting that to some degree code and experimental apparatus are analogous. Since there seems not to be a push to force experimentalists to share their machines, maybe it is unfair to expect computational physicists to do the same without an incentive.

At the other end of the spectrum, Konrad Hinsien argues in his [Correspondence](#) that closed-source software is actually antithetical to the ideals of science and should be banned. The fact that the modelling implemented by the code is hidden from the user, referees and readers of the published paper gives considerable concern for the accuracy and reproducibility of the results.

However, for those of us who wish to engage in best practice for code development and sharing, we are delighted that Radovan Bast from the CodeRefinery project has written a [Comment](#) outlining some guidance on how to achieve this.

One crucial point that we think is worth reinforcing is that the specific action to be taken on these issues depends on the motivation. For example, enabling other researchers to use data in their own work to push knowledge forward more quickly requires different data to be shared than if the motivation is to combat scientific fraud

or to enhance reproducibility. Perhaps part of the issue in the debate thus far is that the link between the motivation for open science and the practical implementation of it has not been made sufficiently strongly.

As a journal, it is crucial for us to reflect on these responses and put into place systems and procedures that encourage and complement the consensus of the research community. Based on this feedback, it is clear to us that some movement on sharing of data is necessary.

From now on, we will strongly encourage the authors of papers that we publish to upload the data that go into the final version of the plots as additional files. For example, if a subfigure contains a line graph with four lines, we would like the values of the data points of those four lines (and their error bars) to be available. Or, if a subfigure contains a two-dimensional colour plot, we would like the data file that contains the x and y coordinates and colour value to be shared.

The motivation for this is simple: although it is common practice for published data to be shared informally if someone asks the authors directly, we feel that having these data easily available on our website speeds this process up substantially and reduces the potential for misunderstanding.

One thing that we wish to stress is that we will not force authors to do this if they do not want to: sharing of these data is not going to be mandated in the foreseeable future. Also, we are also not asking for large files of raw data or the intermediate steps of semi-processed data. Imaging data or conceptual sketches are also not included. In this regard, our current Data Availability policy (<https://go.nature.com/2M3FT3z>) remains unchanged.

Most importantly, we hope that this will be seen as a marker for researchers that we believe sharing data to be a fundamentally good idea. This is an easy and practical step that we can take to help the community move down this path. Hopefully, this will begin to normalize the idea of sharing data in subfields that currently do not think about it much, and be a practical help to researchers at the same time. □

Published online: 1 August 2019
<https://doi.org/10.1038/s41567-019-0641-2>