A problem shared is a problem halved

After much debate about what should be done about sharing of scientific data and source code, practical solutions are still hard to come by. How should the physics community move forward?

In 2009, a small controversy erupted when an independent team of researchers used publicly available data from the Fermi gamma-ray telescope to claim possible evidence for dark matter¹. The sticking point was that the paper written by the Fermi collaboration had not been published yet, so they had been ‘scooped’. Some thought that this was a public relations disaster for the collaboration, but others hailed it as a step forward for open science.

Ten years later, the discussion around the public sharing of data is still going and common practices vary between different subfields of physics. While data sharing is common in astronomy, only a few large collaborations, such as the IceCube Neutrino Observatory² or the Virgo and LIGO Scientific Collaborations, routinely share all of their data, and some might do so only after an embargo period. Small labs frequently do not share their data at all. It is understandable why this could be: careful archiving and curating of data is a time-consuming and expensive process that would be a significant drain on resources for a small lab. But there are also more personal reasons: if a lab is in competition with many other groups who have similar technical capabilities, they may not want to give up any small advantage that they have.

Many large-scale physics facilities, such as the ISIS Neutron and Muon Source in the UK, insist that users make their data publicly available, but some, including the Spallation Neutron Source at Oak Ridge in the USA, merely encourage it. So, consensus on this issue has not even been achieved between facilities that provide similar functions. In fact, the mandate at ISIS comes directly from the Science and Technology Facilities Council, the UK government body that funds the facility, further muddying the waters of who really sets the policy.

Apart from the obvious advantages that open data could bring for reproducibility and scientific integrity — although there is devil in the detail of how to implement this³ — one argument in favour of data sharing is that having more pairs of eyes on data may lead to new discoveries and faster scientific progress. The multiple analyses of LIGO gravitational wave data (sometimes with opposite conclusions) is an obvious

example of this. But within the condensed-matter community, there is also scepticism that such openness may lead to a ‘new’ mode of research where people analyse the data of others without ever doing an experiment themselves.

There can also be more practical considerations that drive this dichotomy. A specialist in complexity might use data from a social media platform to perform network analysis. These data could contain personal information and so there are ethical concerns to take into account. If the data come from a third-party organization, access may only be given to the researcher on the basis that the data are not shared more widely, so there are legal issues as well.

Careful thought is needed about what features are truly deliverable.

Many also advocate the sharing of source code, citing similar advantages for reproducibility and openness, but again there are complications. Licences for commercial packages, such as the VASP density functional theory suite commonly used in condensed-matter physics, explicitly prohibit distribution of their code, a seemingly insurmountable incompatibility with the principle of openness. Also, much home-grown code requires specific environments and library dependencies that are not always trivial to document and manage. The increasingly widespread use of virtual environments may alleviate this to some degree, but adds significant resource constraints.

It is also clear that publishers have a role to play in this fragmented landscape: ideally, consistently enforced shared policies between different journals would help authors and reviewers know what is expected, but careful thought is needed about what features and capabilities are truly deliverable. As one example, several Nature Research journals recently partnered with Code Ocean to conduct a trial using its code review and sharing platform⁴, the results of which should be available soon. This platform creates Docker environments for code and data, and

provides a certain amount of run time for peer reviewers and other users to test the code. The environment is almost certainly not suitable for large-scale computational physics simulations, but could be helpful for smaller data analysis and processing scripts.

At *Nature Physics*, in addition to encouraging researchers to share as much as they feel is feasible, it is now mandatory to include a data availability statement in every published paper⁵. Most of these simply state that the authors will grant access to data “upon reasonable request”, and this is surely a modest step forward. Indeed, in the near future, perhaps it is achievable to insist that all the data used to create the plots in a paper are available with the paper itself. Certainly, this could help theorists who wish to compare their calculations to published experimental data, and vice versa.

So, despite progress in the last ten years, it seems clear that there is no consensus within the physics community on the best way to proceed. At *Nature Physics*, we feel that it would not be desirable to force sweeping new policies on researchers without being sure that this is welcome and constructive. So, we would like to facilitate this conversation and hear a wide range of input from researchers. In the coming months, we hope to be able to publish a few short comments on these issues to highlight areas of consensus and chart a path forwards. What practical suggestions can you make? Are there particular tools or best practices that are helpful? Or perhaps you feel that this push is unwise and want to raise specific unintended consequences? If you wish to voice your thoughts, please e-mail them to naturephysics@nature.com and we will share the most insightful contributions. □

Published online: 1 February 2019
<https://doi.org/10.1038/s41567-019-0434-7>

References

1. Cartwright, J. Dark-matter paper raises questions over data sharing. *Physics World* <https://go.nature.com/2QV2oq7> (29 October 2009).
2. *Nat. Phys.* **15**, 1 (2019).
3. Chen, X. et al. *Nat. Phys.* <https://doi.org/10.1038/s41567-018-0342-2> (2018).
4. *Nat. Methods* **15**, 641 (2018).
5. *Nat. Phys.* **12**, 889 (2016).