

Data sharing varies across physics



Data sharing has been a success for large collaborations like CERN and LIGO. But what about small research groups with bespoke experimental setups?

The physics community is an open one, as can be seen from the success of the arXiv, where preprints have been freely posted and shared with the community for the last 30 years. So, the push for open data from funding agencies in recent years should not come as a challenge for physicists. Indeed, some physics communities, particularly those involving big collaborations such as high-energy physics and astronomy, have successfully embraced data sharing, which has led to new scientific insights. However, the path to meaningful data sharing is not so clear for communities that are formed out of small research groups, like condensed-matter or optical physics. We encourage funders and policymakers to consider the varied needs of the diverse physics community.

In this issue, we publish a [Viewpoint](#) article, in which we asked researchers at various shared facilities about their opinions on open data. The feedback from the particle physics and gravitational wave communities was largely positive. Jonah Kanner, from LIGO told us that about 100 papers have been published a year from researchers not directly involved in LIGO, based on analysis of publicly available data since 2014. Similarly, research-quality open data released by the CMS experiment at CERN since 2014 has led to publications from researchers who are external to the collaboration, says Kati Lassila-Perini, a leader of the open data effort at CMS. The learnings from the CMS initiative have now been adopted by CERN's new [policy for open science](#).

More recently, other shared facilities have followed in the steps of CERN and LIGO and brought in policies to share data after an embargo period of three years. These facilities include the National High Magnetic Field Laboratory (LNCMI) in France, the Synchrotron-Light for Experimental Science and Applications in the Middle East (SESAME) in Jordan and the Extreme Light Infrastructure (ELI) project in Romania. Such experimental facilities are mainly used by condensed-matter physicists, nuclear and materials scientists – fields in which research is typically undertaken by small research groups, rather than big collaborations.

Despite the push for open science mandates at these facilities coming from funders and policy makers, both Andrea Lausi (SESAME) and Charles Simon (LNCMI) commented on the lack of resources, in terms of trained personnel and data storage capabilities to put open data policies into practice in a meaningful way. Simon also highlights the small number of people involved in each technique. Unlike CERN and LIGO, which have a history of

standardized data (crucial for a big collaboration to work), these smaller facilities have been serving a diverse community of scientists, and so have not developed a central, standardized infrastructure that naturally lends itself to data sharing.

Indeed, it is unclear how infrastructure could be developed to facilitate meaningful data sharing for such facilities. Sophia Chen (ELI) has reservations about how this would work at their light source, where users typically bring their home-made diagnostic setups to the facility, making it difficult to standardize data and metadata formats. In many areas of physics, such as condensed-matter, nuclear, or atomic, molecular, and optical physics, the uniqueness of each experiment is the key to their success. The nature of these fields means that small groups are all investigating slightly different research questions, with their own bespoke setups, and these small, sometimes mismatched, pieces of the puzzle are used to build a holistic understanding of a topic. This makes it challenging to implement any type of broad open data policy in such fields. Although researchers may want access to the data behind a published graph to benchmark their own results, it is unlikely that they would (or could) analyse a full raw dataset from a different group in the search for new physics – there are simply too many unknowns in someone else's measurement.

The authors of the Viewpoint make several suggestions of how publishers can support data sharing for researchers at their facilities. Ideas include ensuring that the data used to plot graphs in published papers are available in an open access repository, developing processes to track the citation of datasets and establishing shared-cost cloud solutions to data storage. This conversation builds on [discussions had by our colleagues at *Nature Physics* in 2019](#). As optical physicist, Jacopo Bertolotti, writes, “[One size doesn't fit all](#)”. Open data policies must be crafted with the needs of specific communities in mind, and we call for both deeper discussions between policy-makers, publishers and researchers as well as creative suggestions towards this goal.

Published online: 7 February 2023

“Open data policies must be crafted with the needs of specific communities in mind”