

should handle the data generated in these numerical simulations.

A group of us has, therefore, started to establish the *Encyclopedia of Quantum Geometries*, for now as a collection within the larger Zenodo platform that is funded through CERN, OpenAIRE and the European Union.

To establish the desirable parameters for such a project, we sent out a survey to researchers involved with simulations of numerical data in quantum gravity. This survey gathered 33 answers that came from researchers working on a wide variety of approaches, most numerous (50% of respondents) from those in the loop quantum gravity and spin foam communities, but also many (19% each) from those working on causal set theory and causal dynamical triangulations.

This broad community of researchers requires a variety of storage needs. A majority of users estimated that they generate less than 1 GB of data per year, but a sizeable minority expect that they might deposit more than 100 GB of data per year. Also, the size of the files containing the data was estimated to range from a few kilobytes to several hundreds of gigabytes.

Survey participants also had the possibility to write in answers, and many thoughtful opinions were provided. One question of particular interest is how reuse of data would influence the field in practice. On the one hand, this would open the possibility of finding new effects and testing new hypotheses faster than is currently possible, but on the other hand, the ease of reusing data might lead to fewer independent verifications of results. There is also the risk of misunderstanding data, for example if documentation is not as clear as expected. The level of specialization in the field leads to many different conventions that are not always stated clearly and that might lead to erroneous results if a researcher repeating the analysis is not aware of them.

These are serious risks, however, we still believe that they can be tackled and that the advantages outweigh the risk, and the survey participants widely agreed. The overwhelming majority (88%) were interested in sharing their data alongside published papers in this manner (with a further 9% noting that they would either not have data to share, or that they already share their data using the supplemental

data mechanisms of journals). Equally, 88% said that the existence of open data would be useful for their research in one way or another.

This shows that a repository such as the *Encyclopedia of Quantum Geometries* is a highly desirable feature that could help foster the exchange and adoption of numerical results in the wider quantum gravity community. We thus hope that this encyclopedia will be adopted widely and thus contribute to a more open climate within our community. □

Benjamin Bahr^{1,2}, William J. Cunningham³, Bianca Dittrich³, Lisa Glaser^{4*}, Dustin Lang³, Erik Schnetter⁵ and Sebastian Steinhaus³

¹Erlangen University, Erlangen, Germany. ²Hamburg University, Hamburg, Germany. ³Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada.

⁴University of Vienna, Vienna, Austria. ⁵Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, Canada.

*e-mail: lisa.glaser@univie.ac.at

Published online: 1 August 2019
<https://doi.org/10.1038/s41567-019-0626-1>

Slow and steady

To the Editor — For decades, particle colliders have exposed the fundamental building blocks of nature, most recently the Higgs boson, discovered at the Large Hadron Collider (LHC). In 2014, the Compact Muon Solenoid (CMS) experiment at the LHC took the unprecedented step of making a meaningful fraction of their data public. The CMS Open Data project (<http://opendata.cern.ch/>), now exceeding a petabyte of real and simulated collisions, has spawned several exploratory studies^{1–4}, including our recent search for new particles⁵.

Why ‘unprecedented’? Collider datasets are huge and inherently complex. LHC proton collisions occur every 25 nanoseconds, and reconstructing the collision debris requires synthesizing information from hundreds of millions of readout channels. A filter (the ‘trigger’) discards all but the most interesting collisions, and accounting for its effects and those of the heterogeneous LHC detectors is challenging. The resources required to make such a complex dataset public and usable are substantial, but in short supply.

However, data from the LHC — whose successor is decades away — are priceless for future scientists and must be carefully archived, along with all necessary associated knowledge. As it is archived, the data should be made public, though not immediately. A delay of several years, enough for the experimenters who collected the data to perform thorough analyses, is appropriate; only those who spent years building the experiments have earned quick access. Furthermore, making LHC data ready for public use, with documentation and example code, requires significant funding and time.

But steady publication of LHC data has multiple benefits. First, it encourages prompt archiving, before collective memory fades and knowledge is lost. Second, other scientists can analyse the data while the LHC is still running, testing unconventional strategies and potentially leading to unexpected discoveries, new approaches and fruitful discussions. And third, as a by-product, these scientists can stress test the archiving methods; any deficiencies found are easier to fix now than later.

In this way, public collider data can complement the overall LHC research effort. We, therefore, favour a slow but steady approach to full publication of the LHC experiments’ data; it is in the best interest of particle physics. □

Matthew Strassler¹ and Jesse Thaler²

¹Department of Physics, Harvard University, Cambridge, MA, USA. ²Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA, USA.

e-mail: strassler@physics.harvard.edu; jthaler@mit.edu

Published online: 1 August 2019
<https://doi.org/10.1038/s41567-019-0628-z>

References

- Larkoski, A., Marzani, S., Thaler, J., Tripathy, A. & Xue, W. *Phys. Rev. Lett.* **119**, 132003 (2017).
- Madrado, C. F., Cacha, I. H., Iglesias, L. L. & de Lucas, J. M. Preprint at <https://arxiv.org/abs/1708.07034> (2017).
- Andrews, M., Paulini, M., Gleyzer, S. & Poczos, B. Preprint at <https://arxiv.org/abs/1807.11916> (2018).
- Lester, C. G. & Schott, M. Preprint at <https://arxiv.org/abs/1904.11195> (2019).
- Cesarotti, C., Soreq, Y., Strassler, M. J., Thaler, J. & Xue, W. *Phys. Rev. D* **100**, 015021 (2019).