The University Of Sheffield.

Open Research Case Studies

Robert Shaw, Chemistry

Creating new open-source software in computational chemistry

Postdoctoral researcher Robert Shaw and colleagues in the Department of Chemistry have successfully engaged with open research through the development of open-source software project libecpint.

Computational chemistry is increasingly used to guide and interpret experiments, as well as develop and test underlying theories of how chemistry happens. An important example is the modelling of systems containing heavier elements of the periodic table. These play vital roles in a range of areas, including improving sustainability of the chemical industry, producing new smart materials, and the nuclear fuel cycle. However, the 'effective core potentials' required for accurate and efficient calculations on these systems are typically only available in proprietary software.

Applying open research to computational chemistry

When we looked at creating a piece of open-source software in this area, we knew that it needed to be reusable and reproducible across various software packages. We therefore set about developing an open-source library, using sustainable software development practices, that would provide effective core potential functionality to other programs. Improving the approaches used in these calculations would also greatly reduce computational expense.

The novel algorithms we developed led to speed-ups of up to forty times over existing literature approaches, and we realised our implementation may also be beneficial in commercial computational chemistry packages. With this in mind, we released the code under the MIT open-source licence, allowing code reuse for open-source or proprietary projects. The next step in making the code accessible to the community was online hosting with version control on GitHub. In order to make the project sustainable, we wanted to ensure that others could contribute easily and meaningfully, so we added a number of helpful features. These included documentation for users and developers, a code of conduct and 'architecture statement' for contributors, and continuous integration to help find and correct errors before they caused problems for the project. Some of these valuable additions arose from engaging in the *Journal of Open Source Software (JOSS)* open peer review process.

The software library itself is an open research output and has been assigned a Zenodo DOI[1], making the software citable and helping to attract additional contributors. We have also produced several open access, peer-reviewed articles [2], including the article published in *JOSS* [3], which provides a statement of need for the software and its functionalities.

"I believe research is most valuable and sustainable when it is accessible to everyone, and open research practices are key to achieving that."

Another motivation for ensuring the openness of our research was alignment to the FAIR principles, and we aimed to make the algorithms required for calculations both accessible and reusable. We feel it is particularly important that a 'reference implementation' such as *libecpint* is open, free and meets community standards for sustainable software, as one of its primary purposes is to help in the creation of other software implementations. Our open approach has therefore enabled users to adapt and further develop the work themselves.

Looking to the future

The impact of making this research and the resulting software open is that it has now been incorporated into at least four computational chemistry packages. These include the commercial/free for academic use package Entos and the open-source packages QCSerenity, VOTCA and Psi4. Notably, the inclusion in the popular Psi4 package led to my contributing to, and becoming a named author of, Psi4 and its corresponding journal article [4].

It was surprising, but positive, to find there was a larger demand for the software than had been anticipated. However, this has been something of a double-edged sword; while other interested researchers have contributed code that improves the software, there have also been requests for additional features or changes that have led to extra work. There have also been difficulties in navigating various sustainable software technologies, such as code inspectors and continuous integration, with limited expertise, time and resources.

Robert Shaw, postdoctoral researcher in the Department of Chemistry

Funding or recognition for ensuring that scientific code is open and sustainable has been incredibly limited in the past, and it is very pleasing that the scientific community is making large strides to address the culture of irreducibility. On a personal note, the positive aspects of the time and effort invested are an increase in important skills and the knowledge that the software will be usable, and improvable, by the community for years to come.

Our open research

- Open-source software accessible to contributors for further development
- Publication in open access journal with open peer review
- Software subsequently incorporated into other open-source packages

References

[1] Shaw, R. and Hill, J. (2021). Libecpint. [Software]. Zenodo. https://doi.org/10.5281/zenodo.4694353
[2] Shaw, R. and Hill, J. (2017). Prescreening and efficiency in the evaluation of integrals over *ab initio* effective core potentials. *The Journal of Chemical Physics* 147: 074108. https://doi.org/10.1063/1.4986887
[3] Shaw, R. and Hill, J. (2021). libecpint: A C++ library for the efficient evaluation of integrals over effective core potentials. *Journal of Open Source Software* 6(60): 3039. https://doi.org/10.21105/joss.03039
[4] Smith, D. et al. (2020). PSI4 1.4: Open-source software for high-throughput quantum chemistry. *The Journal of Chemical Physics* 152: 184108.