Our view on open source code development for scientific software at the Center for Space Environment Modeling at the University of Michigan

Gabor Toth
Research Professor
Center for Space Environment Modeling
University of Michigan
gtoth@umich.edu, 734-615-5362

We had multiple discussions of the proposed open source development idea put forward by NASA. Here we briefly summarize the main points that our group agreed on:

1. We believe that reproducibility of research is essential, which means that the various ingredients used should be available for independent checking. These include the data, the methods and software. Having access to source code is essential, as large codes cannot be easily reproduced even if their algorithm is published, and running an executable hides the actual algorithm being used. Our center has been the first in the space physics community to make our source code fully available. This is sufficient for checking reproducibility, but it is not the same as open source access.
2. Most open source code developers do not make a living from this activity. In contrast, graduate students, post docs and research scientists on soft money make a living from various activities that involve, as a major component, developing and maintaining software that can be used for scientific research. Under the current funding practices the only way a model developer can truly benefit from their work is by making some restrictions on the use of their model. For example, our center requires that external users of our code do not directly compete with us for funding. They are expected to collaborate and involve us in their proposals. With no restrictions, the model can be used by anyone who can then get funded for their scientific work that makes use of the model, while they did not actually contribute to the development. This would take away the incentive from putting any serious effort into model development.
3. There are a few examples where people involved in open source development make enough money to sustain their activity for a long period of time. One example is the Geant 4 code developed at CERN. The developers are paid for the code development by CERN as long as there are enough users. Another example is the GNU project, where the developers are paid by various donors through the Free Software Foundation. The donors include the companies who benefit from the GNU software. The current funding model of NASA and NSF is different: people are not (or only partially) paid for software development, and it is assumed to be done as part of a scientific investigation for a limited time and for a limited amount of money. This funding model is incompatible with sustained open source development that pays for a significant (large fraction of total) work effort.
4. The analogy between “observational data” and “source code” is not accurate, and in fact, it is misleading. Observational data is analogous to the output of running a model.

The executable code is analogous to the instrument making the observations. The source code corresponds to the detailed blueprints of the instrument. To the best of our knowledge, no instrument team has ever made their detailed blueprints available, and there is no requirement to do so. Even if the blueprints were available, building the instrument would take significant investment and effort. In contrast, if the source code is available under an open source license, there is negligible effort to make a copy and start using it.

5. The university owns the copyright for source code developed at universities. Even if the code developers would agree to make their code open access, the university may not agree to that. This means that any open access requirement would have to be negotiated with the universities.
6. Open source code development can be advantageous and efficient in some cases. New projects may benefit the most. Open source development could/should be encouraged in general, but making it a requirement, especially for existing software with decades of development, is not a good idea. On the other hand, our center believes that source code should be accessible (for reproducibility), and this could potentially be a requirement for publications and proposals. The current "data policy" of AGU journals has words that require access to "new" source code, but in practice there is no enforcement.

In summary, our center is in favor of public access to source code and we think this should become a requirement eventually. On the other hand, open source code development, while could be encouraged, should not be made a requirement under the current funding models.

One could possibly require making old versions (say after 5 years) of a model available, however, without documentation, maintenance and an active user base, old model versions would serve little purpose other than satisfying a formal requirement. Open source development only works if all participants are willing, motivated and capable volunteers.

Appendix: List of Supporters

1. Yuxi Chen, yuxichen@umich.edu, postdoc
2. Darren De Zeeuw, dezeeuw@umich.edu, research scientist
3. Tamas Gombosi, tamas@umich.edu, professor
4. Kenneth Hansen, kenhan@umich.edu, research scientist
5. Zhenguang Huang, zghuang@umich.edu, postdoc
6. Xianzhe Jia, xzjia@umich.edu, professor
7. Enrico Landi, elandi@umich.edu, professor
8. Michael Liemohn, liemohn@umich.edu, professor
9. Ward Manchester, chipm@umich.edu, research scientist
10. Igor Sokolov, igorsok@umich.edu, research scientist
11. Valeriy Tenishev, vtenishe@umich.edu, research scientist
12. Bart van der Holst, bartvand@umich.edu, research scientist
13. Daniel Welling, dwelling@umich.edu, research scientist
14. Shasha Zou, shashaz@umich.edu, professor

All the above people are affiliated with the Center for Space Environment Modeling at the University of Michigan.