Comments on Best Practices for a Future Open Code Policy for NASA Space Science

J.D. Huba
Naval Research Laboratory
phone: 202-767-6863
email: huba@nrl.navy.mil

1. Positive and negative aspects of NASA-funded science codes required to be open-sourced.

   Positive if the code developer (or developers) are supportive of open-source policies, want the code disseminated for scientific research by other groups, and are willing to provide support to assist users using the code (by providing a tutorial/user's guide to running the code, and addressing issues that may arise from users via, say, email). One significant benefit to the developer(s) is that independent users and researchers can find 'bugs' in the code, which can be corrected to improve the code (i.e., beta testing).

   Certainly negative if the code developer (or developers) are not supportive. It would burden them with additional work they would not want to do, and if forced to do this, they would likely do a minimal or poor job (i.e., actually discourage scientists from using their code by, say, making it very difficult to use).

2. Consequences of requiring existing codes developed by NASA funding be made open source.

   One issue is the definition of an 'existing' code. Codes evolve over time. A code that was used in a published research project several years ago is unlikely to be the same code in use now. It's conceivable (and probably likely) that the latest version of a code would give different results than the one several years ago (hopefully not too different). So, which code should be open-sourced? Both? The newest version?

   Additionally, what if a code was developed several years ago under NASA-funding, but there is no longer NASA-funding for the project. The code still exists. Should the developer open-source it for free?

3. Lessons drawn from personal experience.

   I co-developed (primarily with Glenn Joyce) the NRL ionosphere/plasmasphere models SAMI2 and SAMI3. (The essential difference between these two models is that SAMI2 is a two-dimensional model (latitude/altitude) and SAMI3 is a three-dimensional model (latitude/altitude/longitude). SAMI2 was developed in the late 1990s over a period of roughly 18 months. We were fortunate that a very simplistic ionosphere model had been developed at NRL in the mid-1970s which we had access to and provided a 'blue print' on how to develop a more advanced ionosphere model.

   Circa 2000 I decided to open-source SAMI2. It seemed like a good idea: it would allow other researchers the opportunity to model ionospheric dynamics, and importantly, allow them to modify the code to study various ionospheric issues (e.g., how sensitive is the electron density to different chemical reaction rates which are not well-known).

Both my co-developer and division superintendent were not supportive of this decision. Glenn felt this was a lose-lose proposition, primarily because of a bad experience he had years earlier with allowing other researchers to use a code he had developed. If the code were found to have a bug(s) this could be used as a means to discredit the code developer scientifically. On the other hand, if the code performs well and is used for research then the developer may not get proper (or any) credit. Hence, a lose-lose situation. My division superintendent was concerned about competition. If the code were open-sourced the we would lose our competitive edge in competing for research funds. Despite these concerns, I open-sourced SAMI2 [https://www.nrl.navy.mil/ppd/branches/6790/sami2].

The fears expressed by Glenn and my superintendent were, fortunately, not realized. In my estimation the decision to open-source SAMI2 has been a win-win situation. First, a year or so after the code was released, in the same week, two independent researchers let me know the code 'didn't work.' They were running the code for very low solar activity conditions (e.g.., F10.7 = 70). We had never tested the code for these conditions and were able to fix this 'bug' relatively easily to make the code more robust. Second, to my knowledge, we get proper acknowledgment in published articles that use SAMI2 (this is part of the license agreement specified in the code). However, there was one incident at a CEDAR meeting where a poster described the development of a new ionosphere model that was essentially SAMI2 without any acknowledgment to NRL. They even used a figure from a SAMI2 presentation I had given with no attribution. Needless to say I discussed my concerns with the presenter of the poster - a student who professed naïveté. Third, there have been well over 600 downloads of the code and it seems to have been well-received by the community. A number of researchers have complimented me on it's transparency and ease of use.

On the other hand, I have not yet open-sourced SAMI3. One other reason I open-sourced SAMI2 was that it could be run on a single-processor PC in a reasonable amount of time; thus, it could be used by virtually any scientist or student with a 586 machine in the early 2000s. This was not the case with SAMI3. SAMI3 is parallelized and uses MPI which, until several years ago, required multi-server systems not readily available to many users. The situation has now changed and multi-core computers are available which are adequate to run SAMI3. However, SAMI3 is quite a bit more complex than SAMI2 and it would require a significant effort to develop a user manual for the code which I have not had the time to do (especially unfunded). I have released SAMI3 privately to several research groups for collaborative projects. They have been able to run the code with direct assistance from me and continued dialogues.

To summarize - my experiences with open-sourcing SAMI2 have been very positive - both for me and the science community. That said, to be successful the code developer(s) has to be fully on-board with the process and be willing to spend time with users to assist them.

4. Approach to encourage open-source codes.

   Funding. Doing a good job to provide support (e.g., development of user guides, tutorials, and direct interaction with users) can be a large investment of time and would require adequate funding from NASA.

5. Impact on research if electronic compendium required.

   Probably stifle and slow down research. Is an onerous regulation that would burden the code developer unnecessarily. Moreover, even if the code/inputs/outputs were available, what if the code required 1000 cores on a high performance computer system - is NASA going to provide additional support for any user who wants to run such a code?

6. Other comments

   Why do this? I understand the needs for transparency and reproducibility, but forcing scientists to open-source and maintain their codes seems a bit heavy-handed to say the least. I'm pro open-source but am put off by these proposals. If NASA wants open-source code it should be voluntary and funded through an announcement of opportunity.

```
!     SAMI2-1.00: LICENSE

!     I hereby agree to the following terms governing the use and
!     redistribution of the SAMI2 software release written and
!     developed by J.D. Huba, G. Joyce and M. Swisdak.

!     Redistribution and use in source and binary forms, with or
!     without modification, are permitted provided that (1) source
!     code distributions retain this paragraph in its entirety, (2)
!     distributions including binary code include this paragraph in
!     its entirety in the documentation or other materials provided
!     with the distribution, (3) improvements, additions and upgrades
!     to the software will be provided to NRL Authors in computer
!     readable form, with an unlimited, royalty-free license to use
!     these improvements, additions and upgrades, and the authority to
!     grant unlimited royalty-free sublicenses to these improvements
!     and (4) all published research using this software display the
!     following acknowledgment ``This work uses the SAMI2 ionosphere
!     model written and developed by the Naval Research Laboratory.''

!     Neither the name of NRL or its contributors, nor any entity of
!     the United States Government may be used to endorse or promote
!     products derived from this software, nor does the inclusion of
!     the NRL written and developed software directly or indirectly
!     suggest NRL's or the United States Government's endorsement of
!     this product.


!     THIS SOFTWARE IS PROVIDED ``AS IS'' AND WITHOUT ANY EXPRESS OR
!     IMPLIED WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE IMPLIED
!     WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR
!     PURPOSE.
```