The role of (Open Source) Software in Planetary Science

Corresponding author:

K.-Michael Aye, <michael.aye@colorado.edu>, Laboratory for Atmospheric and Space Physics

Co-Authors:

Chase Million, Founder, Million Concepts

Andrew Annex, <annex@jhu.edu>, Johns Hopkins University

Abstract (350 character limit):

Software has become an essential part of the Planetary Science research enterprise and the ever-increasing volume of data will increase its importance increasingly. We recommend that contributions to open source software should be recognized appropriately and that efforts are undertaken to decrease the entrance burden for a planetary researcher.

[Link to list of endorsers](#)

Introduction

Amid rapid growth of data volume and complexity (multi-spectral/instrument analyses), planetary science is increasingly relying on Open Source Software (OSS). Planetary science is a unique domain of study because it is heavily cross-disciplinary. The richness and diversity of instrumentation and data is reflected in the diversity of software necessary to conduct Planetary Science research. A far from complete list contains

- Mapping software to geo-locate data
- Stereoscopic software to produce topography data from imaging
- GIS software to analyze multiple facets of taken data together,
- Fluid dynamics simulations to study observed erosion and transport phenomena
- Geophysical simulations to model long term trends in planetary body developments and evolution of their surfaces
- Plasma physics
- Spacecraft hardware and flightpath modeling
- "Big" data I/O, storage and retrieval, databases, etc.
- Machine learning and other advanced statistical modeling and methods

While commercial solutions exist for some of these tasks, there are many reasons why OSS is considered better for the scientific process, as for example discussed in the National Academies of Sciences, Engineering, and Medicine 2018 report “Open SourceSoftware Policy Options for NASA Earth and Space Sciences” (National Academies of Sciences, Engineering, and Medicine *et al.*, 2019), where NASA was explicitly recommended to support this trend:

“NASA Science Mission Directorate should explicitly recognize the scientific value of open source software and incentivize its development and support, with the goal that open source science software becomes routine scientific practice.” (1st recommendation of the Discussion section, page 73).

The development and use of open source tools and software for common research tasks increases reproducibility and reduces duplicated research effort. Many of the categories of software written above are too complex for there to be any reasonable expectation that even large research groups should create their own solutions. Mainstream and well-maintained open source solutions exist for some of these classes of problems, especially those with terrestrial commercial applications (e.g. GIS). Other categories of software need to be heavily adapted or wholly recreated for domain-specific applications in planetary science. For example, the data I/O requirements in planetary science are not precisely the same as those in any other field. And plasma physics models for a planet will be much different than those from Earth. Even in the case of GIS software, the tools must often be modified to work correctly for other planets, e.g. to use non-terrestrial geoids; implementations of solutions for issues like these were possible on much

shorter timescales than in the commercial sector because the tools are open source (as with QGIS or GDAL).

Who does it?

Much of OSS development of general use tools in support of planetary science currently occurs unfunded. This is a bad solution; just because OSS is free for users, it does not mean that one should rely on widely used software to be maintained for free. The current reality is that dedicated scientists and engineers spend their free time or however much they can justify on related grant work to develop and maintain software that furthers the open access to and productivity of science. A predictable result of this is that open source tools that others rely on as critical to their research workflows are developed and maintained haphazardly, released with limited opportunity for training or support, developed only to the extent that they meet immediate research needs, and often abandoned or effectively deprecated when the developers' research projects end.

Career-related recommendations based on (Smith *et al.*, 2019)

The precarious professional situation for the scientific programmer was well described in the white paper “Elevating the role of software as a product of the research enterprise” (Smith *et al.*, 2019) for the recent Astro2020 Decadal Survey, which we fully endorse. However, we feel that they could have been formulated more action-able, hence we took them as a template for our career-related recommendations in the following and added in **bold** our specific addition.

1. Recognize software as part of the career path

Software should be recognized in hiring and career development as a key product of modern astronomical research. Software outputs should be considered in all aspects of academic performance appraisals, career applications, and promotion and tenure review cases.

NASA should create research fellowship positions to focus on research software, to prevent people with potential in this area leaving for fields where their talents are more valued.

2. Improve adoption of best practices for software citation

Journals and reviewers should adopt best practices for assuring that software and other science support infrastructure is properly referenced and cited in articles. Referees and other reviewers should be trained to recognize when such acknowledgement is necessary and ask authors to provide that information.

NASA should add a requirement for software citation, where appropriate, for NASA funded projects.

3. Adopt promotion metrics that acknowledge software and other science support

Departments and other members of the community should adopt and use suggested metrics for promotion and tenure reviews of those scientists whose work and contributions involve software and science infrastructure.

NASA should take contributions to OSS into account for career promotion evaluations.

4. Explore partnerships to support data science staff

Universities should explore means of providing support for data-science faculty and staff, perhaps sharing capacity between academic groups or departments internally or partnerships outside the university.

NASA should explore partnerships with universities to support a healthier, faster, and more sustained growth of OSS.

5. Support long-term technical capacity

Funding agencies (**NASA**) should explore longer-term grants aimed at building and supporting professional (non-student) data science capacity. (**see more below**)

6. Treat software milestones as scientific milestones:

A major (community) software milestone or achievement should be seen as having the same value as a major scientific discovery. For example, key developers of astronomy community software should have the opportunity to give plenary talks in major astronomy conferences, not just in, e.g, SciPy conference.

NASA should encourage funding partners like LPI and others to allow software related presentations at major planetary science conferences.

7. Create community prizes for software contributions: Professional astronomical societies (and other organizations/agencies that support astronomical research) should create dedicated prizes and allow for software contributions to be recognized as criteria of merit within existing prizes. **NASA should collaborate with other funding agencies to create prizes for efforts in OSS.**

State of the art for planetary science data related issues

Data readability

Over the history of solar system exploration each mission was, for the longest time, seen as an encapsulated entity, only having to work within itself, the mission team, without future consideration of the outside world. This paradigm started changing with the introduction of the Planetary Data System (PDS), however, space-operational hysteresis is a real thing and has resulted in a multitude of data formats on the PDS, supported by too lax constraints as given by the PDS3 format description.

This resulted in a plethora of insular solutions and the progress of development of a new PDS4 standard. However, a lot of data in the PDS won't be translated into this new format for a long time to come, and until this is the case, many researchers will over and over again face problems in being able to simply read available data for further processing, reducing enormously the efficiency of scientific investments. The end result of this is a vast duplication of effort by the planetary science community and less time allocated to pure scientific inquiry.

Data access

Currently, planetary science data that is stored in the PDS, is stored at several different nodes, based on approximate types of data (images vs in-situ plasma/particle data) and science use cases (atmospheres vs geoscience vs rings and moons). While domain-specific specialization is reasonable for the producers, an increasing amount of research is performed cross-instrument and even across science domains (remote vs in-situ). To access all these different data, the researcher needs to learn different website interfaces, unified programmatic interfaces do not exist and newer PDS-wide data searches do not support filtering by physical parameters. In addition, every new researcher has to develop their own scheme on how to manage data locally after downloading them.

Both “data readability” and “data access” were also topics of a Planetary Data System roadmap study (McNutt *et al.*, 2017), identifying the same issues.

Data analysis pipeline complexities

As alluded to in the introduction, one of the unique challenges of planetary science is the need for a multi-faceted and multi-physics approach for studying phenomena, often creating fragmented insular solutions to sub-problems, because the planetary researcher does not have the capacity to become a specialist in all the subdomains required. Many planetary researchers are required to get familiar with most or many of these very different software environments and all of the cited softwares have a learning curve and are maintained by domain specialists. Trying to use them as an end-user, the researcher is faced with a lack of unified and well documented tutorials, because most OSS software is focusing on functionality versus documentation due to lack of time and/or funds. This leads to continuous reinvention of existing but undocumented solutions, similar to the situation astronomers faced before the community efforts lead to the Astropy project (Muna *et al.*, 2016). The Astropy community was working on unified solutions to the most general astronomy work tasks unfunded for many years despite the library's worldwide use.

We state that **there is a need for a unified general approach software library providing easier entrance into working with planetary data, helping with the aspects of data access, management, reading and processing, and providing easier approaches to standard tasks like observation geometries and insolation (SPICE wrappers), time conversions, and efficient instrument data display (plotting tools).** Many niche solutions exist and while some activities to unite these are ongoing, we are worried that a lack of funding will slow down scientific progress in a time where data taking rates are exploding world-wide.

Sustainable funding

Available funding through NASA for software is hard to find. The PDART program, while officially supporting to “develop or validate software tools”, has a low selection rate for software proposals; the “Data Analysis Programs” (DAPs) of ROSES do not allow projects that focus on software alone. Funding open source projects via 2-3 year grants also has the risk of leaving projects to flounder and deprecate when no continuation funds can be acquired, which is very likely in today’s funding environment.

The newly created software-related programs E.7 and E.8 of NASA ROSES look promising, but are so far not open to solicitation and are currently not well-documented.

Phrases like:

“SMD seeks proposals for the improvement and sustainment of high-value, open source tools, frameworks, and libraries that have made significant impacts to the SMD science Community” (E.7)

and

“Supplemental open source software awards are used to encourage the conversion of legacy software into modern code to be released under a generally-accepted, open source license (e.g., Apache-2, BSD-2-clause, GPL).”(E.8)

seem to indicate that no new projects would be eligible for these programs and any project applying for these funds would have had to be of high impact already, apparently without being previously funded by SMD.

Software-related recommendations

On top of the career-related recommendations motivated by (Smith *et al.*, 2019) we recommend:

1. NASA/SMD should proactively invest into reducing the repeated efforts from constant re-invention of solutions for previously solved problems related to working with planetary data. We recommend the following possible actions to be considered:
 - a. Defining the needs of the community via a special and well-advertised workshop, held initially for example at a future Planetary Data Workshop, but also extended to town halls at more science-focused conferences, so that the collected data becomes more representative.
 - b. Allocate funding for the creation of general-use software tools and packages (like the Astropy project and based on feedback from a)) to address common research tasks like data access, data reading, data management, time conversions, solar system insolation calculations etc, to be defined by polling the community.

- c. Consider funding lines of smaller efforts (<0.1 FTE) but over long periods of time (5-10 years) for the continued maintenance of high-valued research software.¹
 - d. Initiate and support an analysis group that assesses the state of planetary science software, akin to existing NASA analysis groups, being able to provide regular feedback to the planetary science advisory committee (PSAC).
2. R&A Solicitations should actively encourage the budgeting of work time on OSS, including the support for tutorial sessions and workshops for working with planetary data in parallel with scientific conferences.

McNutt, R. L., Jr. *et al.* (2017) ‘The NASA Planetary Data System Roadmap Study for 2017 - 2026’, in, p. IN43E–02. Available at:
<https://ui.adsabs.harvard.edu/abs/2017AGUFMIN43E..02M>.

Muna, D. *et al.* (2016) ‘The Astropy Problem’, *arXiv [astro-ph.IM]*. Available at:
<http://arxiv.org/abs/1610.03159>.

National Academies of Sciences, Engineering, and Medicine *et al.* (2019) *Open Source Software Policy Options for NASA Earth and Space Sciences*. National Academies Press. Available at:
<https://play.google.com/store/books/details?id=ckqCDwAAQBAJ>.

Smith, A. M. *et al.* (2019) ‘Astro2020 APC white paper: Elevating the role of software as a product of the research enterprise’, *arXiv [astro-ph.IM]*. Available at:
<http://arxiv.org/abs/1907.06981>.

¹ Note that the new programs E.7 or E.8 could maybe support this recommendation, but current phrasing does not seem to allow that.