Open Source to Serve Community Science

Primary author: Dr. Arfon Smith, Mission Head, Data Science Mission Office, Space Telescope Science Institute, Baltimore, MD. 410-338-6371, arfon@stsci.edu

Contributing authors

Dr. Kenneth Sembach, Director, Space Telescope Science Institute
Dr. Nancy Levenson, Deputy Director, Space Telescope Science Institute
Dr. Thomas M. Brown, Mission Head, HST, Space Telescope Science Institute
Dr. Marc Postman, Mission Head, Community Missions Office, Space Telescope Science Institute
Dr. Neill Reid, Science Mission Office, Space Telescope Science Institute
Dr. Massimo Stiavelli, Mission Head, JWST, Space Telescope Science Institute
Dr. Roeland van der Marel, Mission Head, WFIRST, Space Telescope Science Institute

Abstract

Founded in 1982, the Space Telescope Science Institute (STScI) is a recognized world leader in science operations of space-based astronomical observatories. This reputation stems from our role as the primary science operations center for the Hubble Space Telescope (HST) and the James Webb Space Telescope (JWST).

In this white paper, we discuss the potential consequences of an 'open code policy' for NASA Space Science in our capacity as a NASA contractor**. Broadly speaking, we believe this would be a positive change** although there may be some challenges in making public some of the software we develop on behalf of NASA to support the scientific community.

Space Telescope Science Institute (STScI) is responsible for the science operations of space-based astronomical observatories. This includes: managing calls for proposals, conducting peer review, planning and scheduling of observations, data processing and analysis, hosting data archives, and providing education and public outreach content.

The majority of work we do at STScI in service of the astronomical community is performed under contract with NASA. A significant amount of the work we do requires us to develop novel software that would be affected by any open code policy within NASA SMD.

In this paper, we respond to three of the questions posed by the committee:

Question 1: What positive and negative impacts would arise for you, your workplace, your NASA funded research, science in general, education, commerce, society, and so on, if all future NASA-funded science code were required to be open source? For example, what maintenance and support issues might arise from open source policies that would not otherwise arise? What relevant experiences have you had with science codes owing to sharing or access constraints? How might negative impacts be mitigated?

Our job is to support the needs of the astronomical community and maximize the scientific returns of NASA-funded space-based observatories. We believe that a mandate from NASA that science code developed by us under our NASA contracts should be open source would have **minimal negative impact and would broadly be a positive development.**

Reuse of data analysis tools & pipelines: Data analysis tools and pipelines are complex software projects that require significant engineering effort, but many of the individual components can be reused by other teams and projects. This reuse is only possible if projects share (and license) their software appropriately. Both the HST[1] and JWST[2] pipelines are open source and available for the community to inspect, modify and reuse. Significant parts of the JWST pipeline have made their way back 'upstream' to community projects such as Astropy[3].

Reproducibility and replicability of scientific results: Transforming raw data from a space-based observatory into science ready data requires complex data reduction pipelines and powerful data analysis tools. Making the source code available for the calibration pipelines and data analysis tools is an important part of ensuring that astronomers have a good understanding of what decisions were made by us when processing mission data. Should these decisions not be appropriate for the needs of an astronomer, publicly available pipelines and analysis tools makes reprocessing data by that astronomer a relatively straightforward process.

---

[1] https://github.com/spacetelescope/hstcal
[2] https://github.com/STScI-JWST/jwst
[3] https://github.com/astropy/astropy

Collaborating with, and supporting the astronomical community: Based on our experiences of releasing open source software, we find that the majority of the NASA-funded scientific software we develop benefits from collaboration with the community: Open source software is likely to receive extra testing, better user feedback, more serious review of the algorithms and correctness of the software[4], contributions to documentation and tutorials, and potentially bug-fixes and new functionality contributed by others.

Available source code is better than no source code: Not all NASA funded software projects are likely to experience (or seek) the benefits of open source. For example, some projects may have no reason to develop community software, and supporting subsequent users of open source software can require significant effort beyond the reach of an individual investigator or small research group. However, there can still be benefits to the community when the software is released as open source because other projects may benefit from the innovations, even if unsupported by the original project. In addition, platforms such as GitHub have recently added features that allow a project to make clear that they are not providing any user support[5].

Question 2: What would be the consequences, positive or negative, if NASA exercised any rights it may have to require that existing codes previously developed under NASA funding be made open source?

Positive consequences: A mandate for releasing existing codes as open source would result in more software being available to the community. It is possible that open source communities (such as Astropy) could incorporate, update and/or improve otherwise unavailable source code and projects.

Negative consequences/challenges: Some of the early stages of our data reduction pipelines, as well as software related to flight operations, are under regulatory control (such as ITAR) and as such could not be easily released. In addition, there are likely security risks in releasing software that is tightly coupled with our operational systems as this could reveal potential attack vectors against our systems.

Question 4: What lessons can be drawn from your experience with open data policies that might help inform future open source policies?

Starting with the HST, NASA pioneered the idea of there being high quality, science ready, open data available in the archives. The availability of archival data means that for HST (and other missions) there are *more* publications published using archival data than for new observations (see Figure 1).

---

[4] http://www.zdnet.com/article/coverity-finds-open-source-software-quality-better-than-proprietary-code/
[5] https://github.com/blog/2460-archiving-repositories

Open access data is at the heart of what we do as a NASA archive. We firmly believe that combining a mandate for open source software with the existing high standards for open data would serve to improve the scientific productivity of NASA's missions.

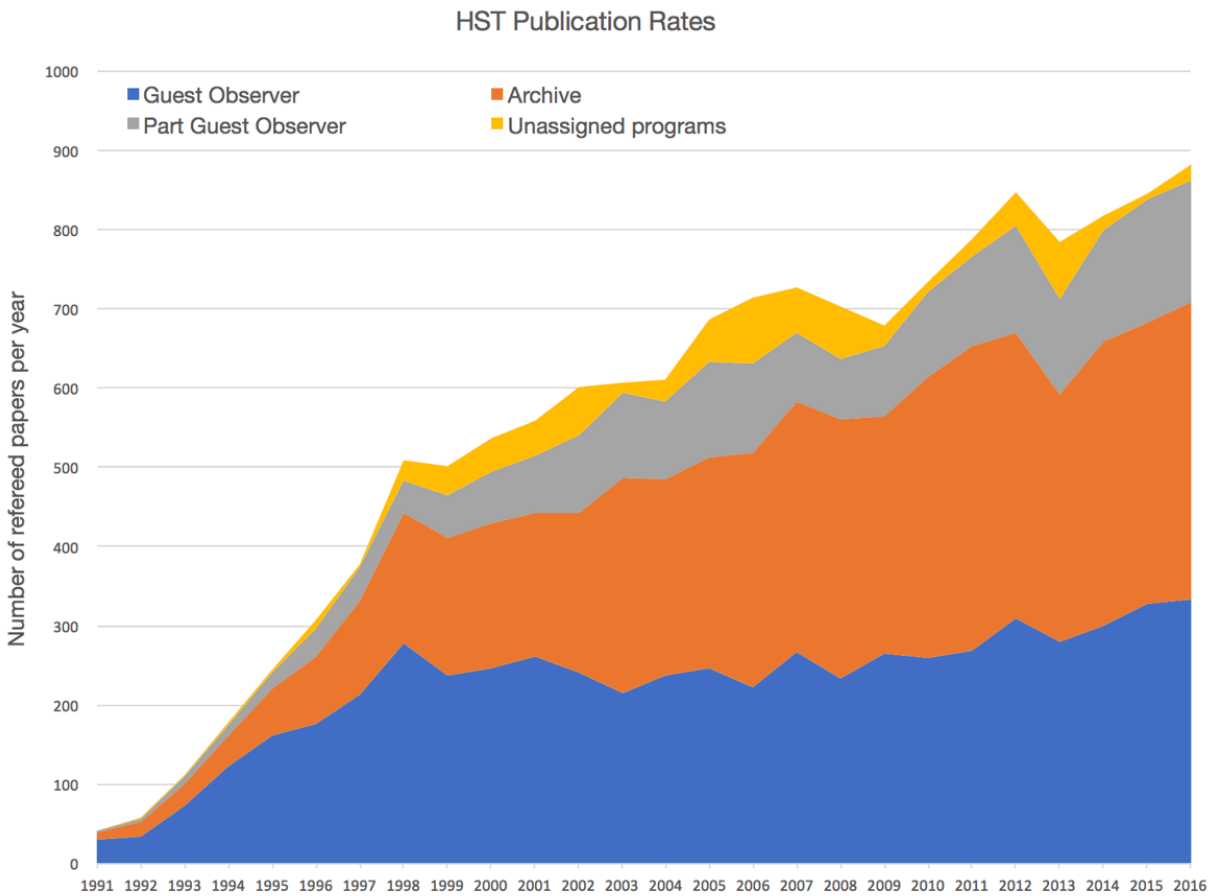

Figure 1: Refereed scientific publications per year since the launch of the Hubble Space Telescope. 'Guest Observer' publications are derived from new observations made by HST, 'Archive' publications are the result of (re)analyses of existing archival (open) data.