Journals of the American Astronomical Society's Response to: RFI for NASA Public Access Plan for Increasing Access to the Results of NASA-Supported Research

Authors Kevin B Marvel, AAS Executive Officer, kevin.marvel@aas.org; Ethan Vishniac, AAS Editor in Chief, ethan.vishniac@aas.org; Julie Steffen, AAS Chief Publishing Officer, julie.steffen@aas.org; August Muench, AAS Data Editor, august.muench@aas.org

AAS Recommendations for NASA

- *NASA should continue to make clear to its investigators that research publication costs in the form of author publication charges are allowable and reasonable costs in any grant proposal and encourage them to earmark these funds for this purpose.*
- *Via its coordination with CHORUS, NASA should seek to make available links to accessible (HTML, ePub) versions of open access articles in their PubSpace / NASA Technical Reports Server, replacing the inaccessible PDF-based model used there.*
- *NASA should recognize that a plan that acknowledges and supports bundling of data and software is key to aiding researchers in making their science open and reproducible.*

Introduction

The American Astronomical Society (AAS) is a 501(c)(3) organization that owns and publishes some of the oldest and broadest leading research journals in the field of astronomy and related disciplines: *The Astronomical Journal* (est. 1849), *The Astrophysical Journal* (est. 1895), *The Astrophysical Journal Letters* (est. 1967), *The Astrophysical Journal Supplements* (est. 1950s), and *The Planetary Science Journal* (est. 2019). *The Planetary Science Journal* was started as a Gold Open Access journal and the others were moved to Gold Open Access effective January 1, 2022. The AAS has never had any of its legacy journal content behind a paywall, and all initially print-only journal content has since been digitized and is hosted by NASA Astrophysics Data System (ADS). The AAS journals collectively include over 5,000 articles per year.

In our role as a non-profit, scholarly publisher that works directly with the astronomy community that comprises our members, we are responding to the Request for Information (RFI): *NASA Public Access Plan for Increasing Access to the Results of NASA-Supported Research*. We provide information on Questions 1, 2, and 5 of the RFI in this document. Our responses to Questions 1 and 2 are directed by our goals to ensure publication opportunities for the community as well as the accessibility and findability of this content. Our investment in high-quality data and software published in the AAS journals through our retention of full-time PhD Data Editors (Muench 2023) leads us to describe nuances and issues on the topic of software sharing and archiving in Question 5.

Question 1: How to best ensure equity in publication opportunities for NASA-supported investigators.

Prior to the move to Gold Open Access, AAS research journals received roughly 2/3 of its financial support from author publication charges and 1/3 from library subscriptions. AAS publication and subscription rates have traditionally been set on a cost-recovery basis and not to recover ongoing operational funds to run the AAS. In making the move to Gold OA, the AAS deliberately declined to use "transformative agreements" or "read and publish" arrangements in large part because our subscription rates were very low to ensure the broadest access to readers and not reflective of author output at subscribing institutions. AAS leadership, in the form of its Board of Trustees, sets all publishing rates for AAS journals and is made up in part of NASA investigators and NASA civil servants who all have the incentive to minimize costs to the community.

Since 2011, the author publication charges for AAS journals have been calculated in "digital quanta," which account for not only the text but the figures, tables, research data, and other digital assets to be published in each article. *The Planetary Science Journal* assigns author charges directly from the number of digital quanta, which ensures that an author's charge reflects the variable amount of actual work necessary to publish the article. The other AAS journals assign charges to authors in three tiers based on the number of digital quanta in the article. To further ensure equity in publishing opportunities, the AAS has also developed an AAS Publication Support Fund; in the case of a shortage of resources, any authors, NASA funded or otherwise, can request assistance from this fund to help cover publication charges in AAS journals.

NASA should continue to make clear to its investigators that research publication costs in the form of author publication charges are allowable and reasonable costs in any grant proposal and encourage them to earmark these funds for this purpose.

Question 2: Steps for improving equity in access and accessibility of publications.

The open access journals owned and published by AAS focus on extensive cooperation between authors, reviewers, stipend-paid subject experts as scientific editors (35), and full-time PhD data editors on staff — all with the goal of helping authors to improve their manuscripts and achieve a high-quality, publishable article. The result is a constructive environment and acceptance rates of 85% or higher in our main journals (50% in the high-impact *Astrophysical Journal Letters*). This ensures that especially early career researchers get the assistance they need to optimize the expression of their research findings and related digital assets and data (Schwarz 2022). The AAS is committed to this community service and emphasizes the communal effort involved.

The AAS has been a pioneer in electronic publishing, starting with the first HTML journal in the field, *The Astrophysical Journal Letters*, in 1995. AAS journals have long been and continue to be produced not only as traditional PDFs, but also in JATS XML, HTML, and ePub (HTML5-based) versions that are more machine-readable than PDFs, which is the preferred format at the core of the NASA PubSpace/ NASA Technical Reports Server. Moreover, the HTML version of AAS journal articles has long been the version of record, containing all the digital assets, including data, published by the author (Kennicutt 2001). HTML-based formats enable critical accessibility tools unavailable in PDFs such as text (font) transformations, screen readers, and reflowing of documents. AAS journal open access content is

carefully edited and deeply structured so that neither NASA nor individual authors need to rework the articles for searchability, easy reading, and preservation in perpetuity.

Via its coordination with CHORUS, NASA should seek to make available links to accessible (HTML, ePub) versions of open access articles in their PubSpace / NASA Technical Reports Server, replacing the inaccessible PDF-based model used there.

Question 5: Suggestions on sharing and archiving of software

Considerations about the sharing and archiving of scientific software (hereafter "research software") are best informed by recognizing that there are different categories of research software created by NASA-supported investigators. As described in the NASA's Public Access Plan for Increasing Access to the Results of Scientific Research, the categories of research software include scripts developed in the analysis of data, compilations of modules or a singular tool closely tied to a particular project, or packages intended to support missions. Such categories vary for instance in the degree of maintenance required (low, medium, high; Martinez-Ortiz et al. 2023).

Since the adoption of our software policy in 2016 (Vishniac & Lintott 2016), the AAS Journals have worked with authors to enhance research software's role as a first-class research object in scholarly literature. This experience includes improving software citations and their indexing (Henneken et al. 2022), working with authors to cite and archive software related to published articles (Muench 2023), and enabling simultaneous science and software peer-review (Vishniac & Lintott 2018). *Based on our experiences working on software with authors, we write to point out that there are aspects of the first and second categories of research software that may require unique forms of support for sharing, archiving, and reuse.* In these categories, software and data are often bundled together. One reason for such bundling is to achieve reproducibility; another reason is that the software provides interoperable access to the data. Although such bundles may be understood to fall under the low-maintenance category of "scripts", data+software bundles have some important characteristics that need to be supported when they are shared and archived.

First, software provides critical context or description of the data in many cases. Such "description" is akin to metadata about the data, and the results may only be reusable with the software created to manipulate it. This will be especially true when datasets have no external curation that ensures that the resulting data are encoded with standard metadata. Simply unbundling data and software when it is archived may leave the data unusable. For reproducibility bundles, software additionally encodes various algorithmic dependencies between data and the specific results. Segregating data from software in these bundles eliminates the transparency desired from open science mandates.

Second, these bundles may contain data reused in a computation as well as data created by the software. For example, a bundle may contain a comma-separated values (CSV) file of source data downloaded from a particular journal article as well as a CSV file of new results. Or the source data may be derived from an application protocol interface (API) to a NASA or Virtual Observatory (VO) database. Software may encode the provenance of the former, and it definitively defines the origin of the latter source. Ensuring the provenance of these data sources is not simply a matter of adding dataset DOIs to the metadata of a bundle – this is foremost because dataset DOIs ***cannot*** provide machine-readable

access to data. Further, many NASA databases are unversioned, nor do they provide a "state" to ensure the results of a particular query. A download of data about known planets in the universe from the Planetary Systems Table [doi:10.26133/NEA12] of the NASA Exoplanet Archive is valid only at the ***specific moment of the software query*** as this NASA database is unversioned. To put that another way, the reproducibility of a result is often encoded **only** in the software or in the source data preserved in the bundle created by the researcher.

Finally, software may be in fact required to access and reuse the data. For example, the project "gPhoton", hosted at MAST, is both a database of time-tagged photons from NASA's GALEX mission and the software used to manipulate or convert the results in that database into new results (Million et al. 2016; Million & Flemming 2016). Notwithstanding the aims of the Virtual Observatory, which include standardizing software services for data, "software" may in many cases be so closely bound to a particular dataset that attempts to disaggregate data from software as a matter of standardization may in fact reduce the efficiency of reuse for NASA-supported investigators compared to the functionality of specialized software tuned to that dataset.

As described above, software and data often and at many levels, from reproducibility bundles to major project results, cannot be simply divorced. *The examples and details we've described for data and software bundles point to unique considerations when NASA guides researchers on sharing and archiving these materials.* Some of the unique considerations include mixed licensing for data+software reuse; resources to support reproducible computation; resources to support reproducible API access to databases and external resources; guidelines on packaging data+software bundles. More specific examples of these unique considerations include: guidelines on combinations of CC0/Public Domain licensing for data and permissive MIT licensing for software; funding reproducible computation with tools such as `binder` (https://mybinder.org/) or the Timeseries Integrated Knowledge Engine (TIKE) JupyterHub at STScI; support for versioned or stateful, i.e., persistent results from major NASA archives; creating or adopting guidelines for bundled data+software research objects, e.g., RO-Crate (Soiland-Reyes et al. 2022), e.g., research compendia (Nüst et al. 2017). There are complicated dependencies between software- and data-management for NASA investigators.

NASA should recognize that a plan that acknowledges and supports bundling of data and software is key to aiding researchers in making their science open and reproducible.

References

Henneken, E. A., Blanco-Cuaresma, S., Accomazzi, A., Muench, A., Ioannidis, A., Nielsen, L. H., & Steffen, J. (2022). Asclepias: Software Citations Enter the Scholarly Literature World. In Vol. 54, Issue 1. American Astronomical Society. https://doi.org/10.3847/25c2cfeb.2906303c

Kennicutt, Jr., R. C. (2001). Editorial: Electronic Supplemental Materials for the Astrophysical Journal. In The Astrophysical Journal (Vol. 557, Issue 1, pp. 1–1). American Astronomical Society. https://doi.org/10.1086/323678

Martinez-Ortiz, C., Martinez Lavanchy, P., Sesink, L., Olivier, B. G., Meakin, J., de Jong, M., & Cruz, M. (2023). Practical guide to Software Management Plans (Version 1.1). Zenodo. https://doi.org/10.5281/ZENODO.7589725

Million, C., Fleming, S. W., Shiao, B., Seibert, M., Loyd, P., Tucker, M., Smith, M., Thompson, R., & White, R. L. (2016). gPhoton: THE GALEX PHOTON DATA ARCHIVE. In The Astrophysical Journal (Vol. 833, Issue 2, p. 292). American Astronomical Society. https://doi.org/10.3847/1538-4357/833/2/292

Million, C., & Fleming, S. (2016). A Time-Tagged Database Of Every GALEX Photon Event ("gPhoton") [dataset]. STScI/MAST. https://doi.org/10.17909/T9CC7G

Muench, A. (2023). The Roles of Data Editors in Astronomy. In Science Editor. Council of Science Editors. https://doi.org/10.36591/se-d-4601-04

NASA Exoplanet Science Institute. (2020). Planetary Systems Table [dataset], Last Accessed: 2023-08-14, IPAC. https://doi.org/10.26133/NEA12

Nüst, D., Konkol, M., Pebesma, E., Kray, C., Schutzeichel, M., Przibytzin, H., & Lorenz, J. (2017). Opening the Publication Process with Executable Research Compendia. In D-Lib Magazine (Vol. 23, Issue 1/2). CNRI Acct. https://doi.org/10.1045/january2017-nuest

Schwarz, G. J. (2022). Authors: Improve Your Bibliometrics with Digital Enhancements. Bulletin of the AAS, 54(1). https://doi.org/10.3847/25c2cfeb.07e76602

Soiland-Reyes, S., Sefton, P., Crosas, M., Castro, L. J., Coppens, F., Fernández, J. M., Garijo, D., Grüning, B., La Rosa, M., Leo, S., Ó Carragáin, E., Portier, M., Trisovic, A., RO-Crate Community, Groth, P., & Goble, C. (2022). Packaging research artefacts with RO-Crate. In S. Peroni (Ed.), Data Science (Vol. 5, Issue 2, pp. 97–138). IOS Press. https://doi.org/10.3233/ds-210053

Vishniac, E. T., & Lintott, C. (2016). Editorial: The AAS Journals Corridor for Instrumentation, Software, Laboratory Astrophysics, and Data. In The Astronomical Journal (Vol. 151, Issue 2, p. 21). American Astronomical Society. https://doi.org/10.3847/0004-6256/151/2/21

Vishniac, E. T., & Lintott, C. (2018). Editorial: A Cooperative Agreement with the Journal of Open Source Software. In The Astrophysical Journal (Vol. 869, Issue 2, p. 156). American Astronomical Society. https://doi.org/10.3847/1538-4357/aaf876