

Request for Information: NASA Public Access Plan for Increasing Access to the Results of NASA-Supported Research

Response from Woods Hole Oceanographic Institution (WHOI)

Coordinated by WHOI's Ocean Informatics Working Group

2023-08-14

Woods Hole Oceanographic Institution (WHOI) is an independent organization with over 1000 staff and students dedicated to advancing knowledge of the ocean and its connection with the Earth system through a sustained commitment to excellence in science, engineering, and education, and to the application of this knowledge to problems facing society. A large proportion of our research is funded by federal grants and contracts. There are principal investigators (PIs) in all of our academic departments who receive funding from NASA. NASA's Public Access Plan for Increasing Access to the Results of NASA-Supported Research will be directly relevant to these NASA-funded projects.

WHOI's Ocean Informatics Working Group has been following NASA's Transform to Open Science (TOPS) initiative, and we are pleased to have this opportunity to respond to the Request for Information (RFI) for NASA's Public Access Plan. Comments were initiated by the Working Group, iterated with research data librarians at WHOI's Data Library & Archives, and then iterated with PIs from all academic departments. The expertise of WHOI's research data librarians is relevant to all 5 of the questions numbered in the RFI; the MBLWHOI Library pioneered the Woods Hole Open Access Server (WHOAS) as an institutional repository to preserve and redistribute the intellectual output of the Woods Hole scientific community in digital form.

Overall, we see NASA's Public Access Plan as being relatively consistent with analogous plans from other grant-making agencies with regard to Digital Scientific Data and Peer-Reviewed Publications (PARTS A and B, respectively); however, NASA's Plan is more inclusive with regard to Software (PART C) and more exclusive with regard to physical samples. With regard to this phrase from the Revisions - "Other data products beyond peer-reviewed publications and software" (on p. 2 of the Plan version February 21, 2023) - our assumption is that this refers to digital unclassified research data (in PART A 2.0 Scope) collected by the project but not used in peer-reviewed publications. However, it would be beneficial to clarify whether these "other data products" extend to research artifacts more generally [e.g., consult with Earth Science Information Partners (ESIP) cluster for Research Artifact Citation].

With regard to the 5 questions in the RFI:

- 1. How to best ensure equity in publication opportunities for NASA-supported investigators.*
- 2. Methods for monitoring evolving costs and impacts on affected communities.*
- 3. Methods for monitoring evolving costs and impacts on affected communities.*

We grouped these two questions and highlight the word - “reasonable” - in the following sentence from the RFI: “NASA policy allows supported researchers to charge reasonable publishing costs against their awards.” In general, Article Processing/Publication Charges (APCs) are higher for open access articles in peer-reviewed journals. We recommend that NASA:

- provide cost guidelines for APCs in an effort to help control costs imposed by publishers;
- consider providing cost guidelines or expectations for data management and publishing for awardees (for example, the European Commission expects about 5% of research expenditure to be devoted toward properly managing and stewarding data (European Commission, Directorate-General for Research and Innovation, Realising the European open science cloud : first report and recommendations of the Commission high level expert group on the European open science cloud, Publications Office, 2016, <https://data.europa.eu/doi/10.2777/940154>);
- create proactive methods of monitoring to measure the impacts of these costs on affected communities. These impacts may be included in project reports and/or as a data collection effort by the agency and research communities involved.

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

NASA has already taken great steps towards improving access to publications through the NASA Technical Report Service (NTRS) (formerly PubSpace) public access repository with Clearinghouse for the Open Research of the United States (CHORUS) Publisher Members. However, when a publication is not with a CHORUS Publisher Member, we recommend the following to assist external NASA grantees:

- PARTS A and C (for data and software, respectively) in the Plan include a subsection for “Guidance and Training,” and we recommend also including a guidance subsection for PART B. This would be for guidance in general with regard to article licensing and rights;
- In the submission form (<https://sti.nasa.gov/sti-contact-form/?RequestType=PublicAccess>) guidance could be provided specifically to the DOI for the peer-reviewed publication (i.e., the PI would enter a DOI and the form could return specifications for the allowable file upload).

4. Input on considerations to increase findability and transparency of research.

With regard to findability for data, publications, and software within our scientific domain we recommend that NASA:

- consider adding guidance towards the selection of keywords from controlled vocabularies, for example the Global Change Master Directory (GCMD) keywords. The GCMD keywords are increasingly inclusive of the Ocean sciences domain, and the (new) EarthData Forum enables keyword requests for further expansion.

The RFI asks about experiences institutions and researchers have had with adoption of different identifiers:

- For our institutional identifier, WHOI uses the public Research Organization Registry (ROR). For Digital Science products (e.g., Altmetric), WHOI uses the Global Research Identifier Database (GRID).
- For researcher identifiers, Open Researcher and Contributor ID (ORCID) is used widely by publishers and funders; this has increased adoption among researchers. If a future version of the Plan includes researcher identifiers, we recommend including “Guidance and Training” (e.g., for managing ORCIDs).

5. Suggestions on sharing and archiving of software.

NASA would be setting a precedent by requiring a separate Software Management Plan (SMP) in research proposals. To reduce the number of separate items to submit, we recommend:

- a single template like NASA's new [ROSES-2023 "Open Science and Data Management Plan \(OSDMP\)"](#) to address all together how data, publications, and software will be made available. We note that the ROSES-2023 template is available through the popularly used DMPTool (https://dmptool.org/public_templates). We also note that journals popular in Earth and ocean sciences often combine author guidelines for the availability of data and software, for example AGU journals (<https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Data-and-Software-for-Authors>).

With regard to sharing and maintaining software, we agree that cloud version control (e.g., git) hosting solutions are of great benefit; for example, GitHub is well-adopted in our community for sharing code. The GitHub to Zenodo archiving solution is easy to use and issues a DOI for the archive. However, Zenodo has very few requirements (and lacks guidance) for metadata.

Presently, the PART C (software) "Guidance and Training" subsection indicates that specific guidance will be provided per solicitation for software licensing and metadata to improve findability. NASA might consider here some guidance:

- for metadata to improve findability, such as discovery keywords (see GCMD above) and award numbers;
- for additional metadata to include with code; some repositories (e.g., the Environmental Data Initiative repository (Gries, et al. 2021. <https://doi.org/10.6073/pasta/9d4c803578c3fbc45fc23f13124d052>) and research coordination networks (e.g., NSF EarthCube <https://modeldatarcn.github.io/>) in our domain are providing detailed guidance;
- for software licensing, to distinguish from licensing for data or publications.

Beyond the 5 questions in the RFI, we would like to provide a few additional comments:

With regard to PART A 3.0 requirement "Unique digital object identifiers (DOIs) must be assigned to all datasets supporting peer reviewed publications":

- Some repositories in our domain do not necessarily mint DOIs, but may use other Persistent Identifiers (PIDs). We recommend adjusting this requirement to include community or data-type-accepted PIDs to mitigate the concern that an author might choose to archive data at a general-purpose repository over a domain repository just because of a DOI requirement.

With regard to PART A "Guidance and Training":

- We also recommend general guidance for data licensing, to distinguish from licensing for publications or software.

With regard to our overall comment above about excluding physical samples:

- In the Ocean sciences domain it can be important to assign unique identifiers for physical samples. In particular in the marine geological sciences, International Generic Sample Numbers (IGSNs) are commonly used.
- The inclusion of physical samples as part of open science is a session at 2023 American Geophysical Union Meeting "IN024 - Enabling Open and Interoperable Access to Physical Samples and their Associated Digital Artifacts."