OCSDO
OFFICE OF THE CHIEF SCIENCE DATA OFFICER

MONITORING IMPLEMENTATION OF THE SCIENTIFIC INFORMATION POLICY FOR THE NASA SCIENCE MISSION DIRECTORATE (SPD-41A)

METRICS REPORT

Metrics on 2023 Publications, Data, and Software associated with NASA Research Opportunities for Earth and Space Science Awards

Office of the Chief Science Data Officer

NASA Science Mission Directorate

June 3, 2025

NASA/TP-2025005826

TABLE OF CONTENTS

EXECUTIVE SUMMARY 1

INTRODUCTION 2

EVALUATION OF PUBLICATIONS 3

Publication Metrics 4

Publication Availability 4

EVALUATION OF SCIENTIFIC DATA AND SOFTWARE UNDERLYING PUBLICATIONS 6

Data and Software Metrics 6

Data and Software Availability 7

DISCUSSION 10

Limitations 10

Factors associated with compliant data and software sharing practices 11

Future work 11

AUTHOR STATEMENT 12

SUPPORTING MATERIALS 12

REFERENCES 12

APPENDIX I: GLOSSARY 14

APPENDIX II: RUBRICS FOR DATA AND SOFTWARE AVAILABILITY SCORES 15

Rubric for Data Availability 15

Rubric for Software Availability 16



EXECUTIVE SUMMARY

The NASA Science Information Policy for the Science Mission Directorate (SPD-41a) provides requirements for how scientific information produced from SMD funded scientific activities must be shared. SPD-41a requirements for research awards were incorporated into SMD's Research Opportunities in Space and Earth Science (ROSES) starting with the ROSES-2023 solicitation. This report evaluates scientific publications from 2023 with respect to the SPD-41a requirements to be used as a baseline for future analysis on policy implementation.

Two sets of metrics are presented in this report. The first describes the number of NASA-funded publications that are made openly available. The second provides estimates for the accessibility of scientific data and software that underlie SMD-funded publications. In summary:

1. The open accessibility of NASA-funded publications identified in two databases ranges from 48% to 83%, reflecting differences in detection of NASA-funded publications and varying methods of making publications openly available.

2. Statistical analysis of a representative sample of peer-reviewed publications across each division of SMD indicated that data availability ranged from 20% to 80% and software availability ranged from 6% to 43%.

These metrics provide insight into current successes and challenges for the adoption of open science practices across SMD research, which can be used to inform future policy implementation priorities.



INTRODUCTION

The NASA Scientific Information Policy for the Science Mission Directorate (SMD) (SPD-41a), updated in 2022, provides guidance for how data, software, and publications produced from SMD-funded scientific activities should be made openly available. SPD-41a requirements for SMD-funded research and analysis (R&A) activities were incorporated into SMD's Research Opportunities in Space and Earth Science (ROSES) funding opportunities starting with the ROSES-2023 solicitation. With these changes, most ROSES funding opportunities from 2023 onward require proposals to include an Open Science and Data Management Plan (OSDMP) that describes how proposers will meet SPD-41a requirements, including openly sharing scientific publications, data, and software.

Metrics were developed in 2024 to assess the implementation of SMD's updated scientific information requirements across ROSES-funded activities. These metrics are intended to provide insight into implementation progress,inform future implementation priorities, and support broader assessments of open science adoption and impact across SMD. .

The evaluation includes SMD-funded research outputs published in 2023 and is intended to serve as a baseline for continued monitoring of SPD-41a implementation in the future. Because SPD-41a requirements were first incorporated into ROSES in 2023, the scientific information evaluated in this study predates SPD-41a requirements and publications were not expected to be fully compliant. However, NASA already had some scientific information requirements in place before SPD-41a (e.g., see SPD-41 for a consolidation of federal and NASA policies applicable to SMD), and some scientific disciplines were already implementing many of these practices.

Plans for this study are described in a preregistered report that was completed prior to data collection. This study spanned R&A activities sponsored by SMD's five science divisions: Earth Science (ESD), Planetary Science (PSD), Heliophysics (HPD), Astrophysics (APD), and Biological and Physical Sciences (BPS). Definitions used in this report are consistent with SPD-41a and summarized in Appendix I.

EVALUATION OF PUBLICATIONS

SPD-41a requires that scientific publications resulting from SMD-funded scientific activities be made publicly available at the time of publication. Investigators may meet this requirement by publishing in a journal that makes the publication openly available at the time of publication or by depositing the as-accepted manuscript into NASA PubSpace. Metrics were derived based on data from two services that support public access and discoverability of NASA-funded publications: 1. Clearinghouse for the Open Research of the United States (CHORUS) and NASA Science Explorer (SciX).

CHORUS data were obtained through the CHORUS Dashboard, which was accessed using institutional credentials by authorized users. The dashboard provides analytical capabilities for NASA as a participating funding agency. Publication metadata were exported in CSV format, including fields such as Digital Object Identifier (DOI), title, authors, journal, publication date, funding acknowledgments, and compliance status. Data extraction was filtered to include only publications from calendar year 2023 with NASA funding acknowledgments. NASA-funded research is identified by CHORUS through publisher-supplied metadata and Crossref funding registries, which may present limitations in coverage for publications where funding sources are inadequately reported.

SciX data were acquired through the query interface of the NASA Science Explorer system. Publication metadata are aggregated by SciX from multiple sources including NASA PubSpace, publisher Application Programming Interfaces (APIs), and grant reporting systems. The extracted dataset contained publication metadata with NASA grant identifiers, division codes, and compliance status indicators. Data were exported in structured JavaScript Object Notation (JSON) format with standardized field names following the National Information Standards Organization (NISO) Journal Article Tag Suite (JATS) schema, a standard XML format for marking up scholarly journal articles. While comprehensive coverage of NASA-funded publications is offered by SciX, its identification mechanisms depend on accurate grant reporting and may exhibit different coverage patterns than CHORUS, particularly for collaborative

international research. Normalization of author names and institutional affiliations was required for both datasets to ensure accurate analysis.

NASA-funded publications in both data sets were determined based on funding information specifically mentioned in the manuscripts (e.g. award identifiers). Due to the information that is easily retrievable, the analysis is performed on NASA-funded publications rather than only on SMD-funded publications. Future analysis may focus on SMD-funded publications.

Publication Metrics

1. **Number of NASA-funded publications identified by CHORUS and SciX**: The total number of publications identified as relevant to SMD funding was determined by aggregating data from both CHORUS and SciX. Publications were identified based on funding information included within the manuscripts, such as funder IDs and grant identifiers. The combined dataset was used to establish the overall scope of publications associated with NASA funding.

2. **Number of NASA-funded publications that are openly available**: The number of publications that were freely accessible at the time of publication was determined. This included publications published under an Open Access license and those for which an accepted manuscript was made available through repositories such as NASA PubSpace. This metric aimed to quantify the immediate availability of research outputs to the public.

Publication Availability

In June 2024, the CHORUS database and NASA Science Explorer (SciX) were queried for peer reviewed publications that were supported by NASA awards.

The CHORUS database reported 4,710 NASA-funded papers published in 2023 for the same year, with varying levels of access. The CHORUS data also reported on which papers were identified as open access. Of the 4,710 publications, 48.1% were verified as open access (OA; Figure 1) as reported by the publisher.

The SciX database reported a total number of NASA peer-reviewed publications as 4642 in 2023. A total of 3,875 refereed NASA-supported papers published in 2023 were identified in the NASA Science Explorer (SciX) database as open access. For those in SciX, 83% of publications are reported as open access. This would include publications made openly available by the publisher and as preprints.

This figure serves as a baseline for monitoring the open access of NASA publications. This data presents two methods using existing services for evaluating the open access nature of NASA publications. Preliminary evaluation indicates there may be significant incompleteness when comparing the results from the two services, but further research is required to confirm.

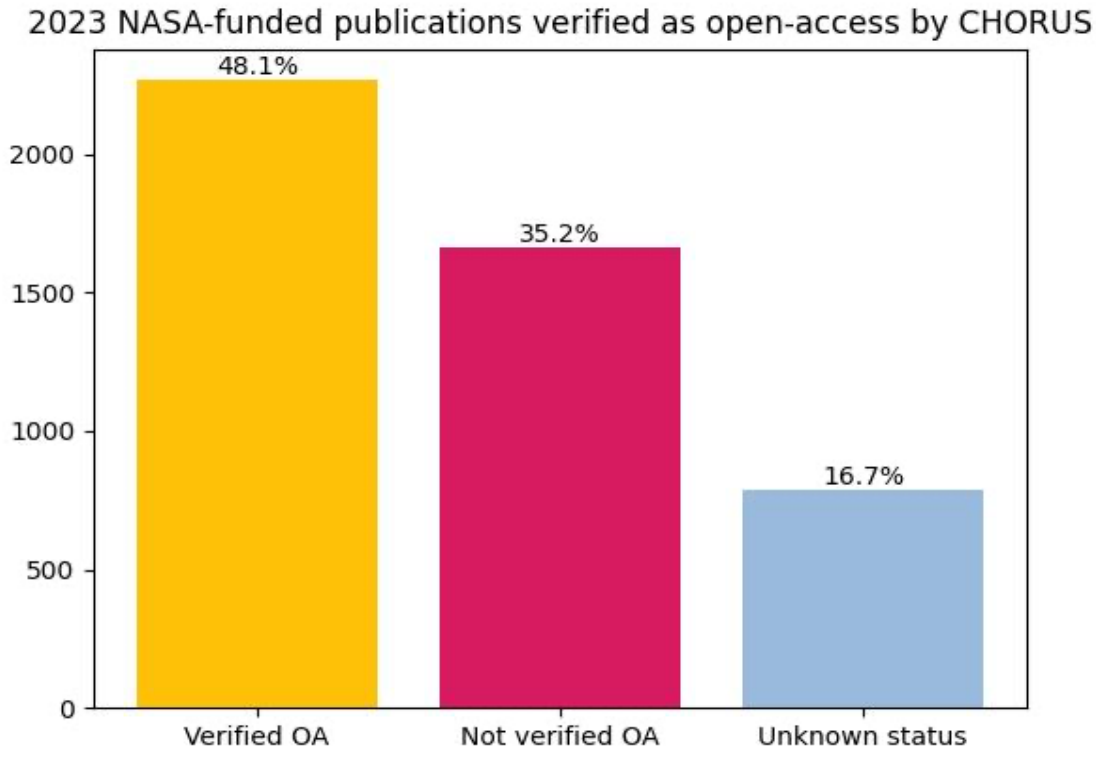

Figure 1: Percent of NASA-funded journal articles verified as open access (OA), where blue is unknown, gold represents the percentage of articles verified as OA, and crimson represents articles not verified OA.

EVALUATION OF SCIENTIFIC DATA AND SOFTWARE UNDERLYING PUBLICATIONS

The NASA Science Mission Directorate Policy Document 41a (SPD-41a) mandates two levels of public access requirements for research products. First, all scientific data and software necessary for the validation of published conclusions must be made openly accessible concurrent with publication. Second, any additional data and software with scientific utility generated during the funded research must be made openly accessible no later than the conclusion of the award period of performance.

Currently, SMD lacks a comprehensive inventory of scientific data and software products generated through ROSES-funded research activities. To evaluate compliance with SPD-41a mandates, we conducted an analysis of publications from 2023 identified as SMD-funded through SciX.

A stratified random sampling protocol was implemented to identify 150 SMD-funded publications, with 30 publications each of the five SMD science divisions. SMD funding status was verified through examination of award identifiers included in publications, funding acknowledgment statements, and cross-referencing with divisional documentation of ROSES-funded publications. Based on the sample size of the population, the estimated error on each measurement is approximately 8%

Data and Software Metrics

1. **Scores for overall quality of data sharing practices described in sampled publications, by SMD division**: Each publication within the sample was evaluated using a standardized rubric (Appendix II) designed to score the quality of data sharing practices. This rubric assessed the accessibility of the data, the format of the data, the presence of a license, and the use of a persistent identifier. Scores were assigned based on the level of compliance with SPD-41a requirements and best practices. These

scores were aggregated and analyzed by SMD division to provide an overview of data sharing practices within each field.

2. **Percentage of sampled publications that demonstrate compliant practices for sharing underlying data, by SMD division**: The percentage of publications within each division that met the criteria for "Compliant" or "Exceeds Requirements" according to the data availability rubric was calculated. This metric provided a measure of the proportion of publications with adequate data sharing practices as defined by SPD-41a.

3. **Scores for overall quality of software sharing practices described in sampled publications, by SMD division**: Parallel to the data evaluation, each publication was assessed using a separate rubric (Appendix II) focused on software availability. This rubric evaluated factors such as the accessibility of the software, the presence of documentation, and the ability to replicate results using the provided software. Scores were assigned based on compliance with SPD-41a requirements and best practices. These scores were aggregated and analyzed by SMD division to assess the quality of software sharing practices across different fields.

4. **Percentage of sampled publications that demonstrate compliant practices for sharing underlying software, by SMD division**: The percentage of publications within each division that met the criteria for "Compliant" or "Exceeds Requirements" according to the software availability rubric was calculated. This metric provided a measure of the proportion of publications with adequate software sharing practices as defined by SPD-41a.

Data and Software Availability

Compliance rates for data and software availability were calculated for the sampled papers. Papers that are considered compliant in this calculation scored 0 (compliant or not applicable) or a 1 (exceeds requirements). Details for individual SMD divisions are presented in Table 1. A graphical representation of these results is presented in Figures 1 and 2.

Division	Data (% Compliant)	Software (% Compliant
APD	20	13.3
BPS	40	13.3
ESD	80	40.3
HPD	20	6.7
PSD	36.7	20

Table 1: Percentage of sampled publications with underlying data and software shared in compliance with SPD-41a requirements. The uncertainty on each measurement is approximately 8%.

Figures 2 and 3 show the normalized data and software accessibility scores respectively ranging from -2 (no effort towards compliance) to 1 (exceeds compliance), with 0 indicating full compliance. Overall, the analysis revealed that publications were more frequently compliant with data sharing requirements than software sharing requirements. Notably, the Earth Science Division (ESD) demonstrated the highest levels of compliance across both data and software sharing.

Further analysis highlighted variations in compliance across different science divisions. While ESD consistently exhibited strong adherence to open data and software practices, all divisions showed publications that were compliant with the requirements in SPD-41a. Except in Heliophysics, most of the publications showed partial compliance with either the data or software sharing requirements. These findings suggest that progress is being made toward open science practices across all the divisions with some publications already being fully compliant. However, there are still inconsistencies in implementation across the SMD divisions. The disparity between data and software sharing compliance indicates a need for targeted strategies to encourage and facilitate the open release of software alongside data.

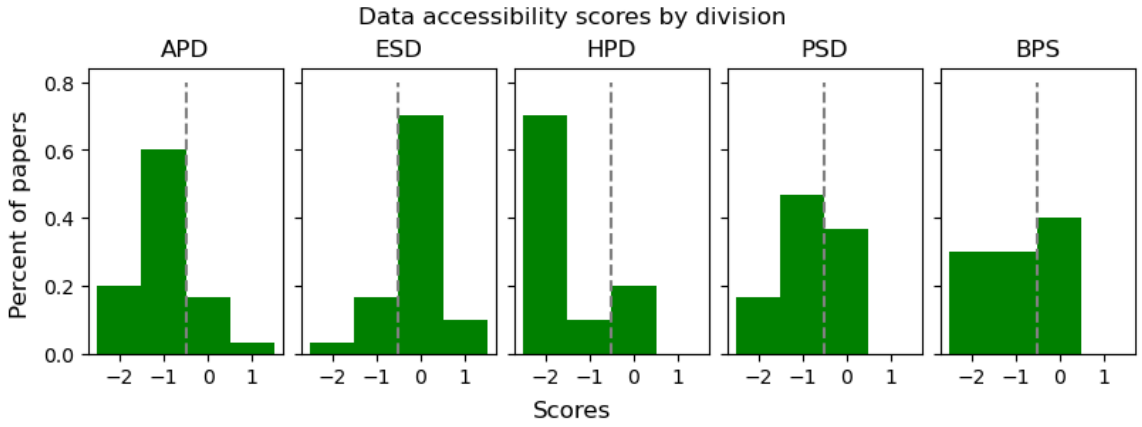

Figure 2: Data availability scores by division. Dotted lines signify the minimum requirements to meet compliance, with papers describing noncompliant practices to the left of the line and papers describing compliant practices to the right of the line.

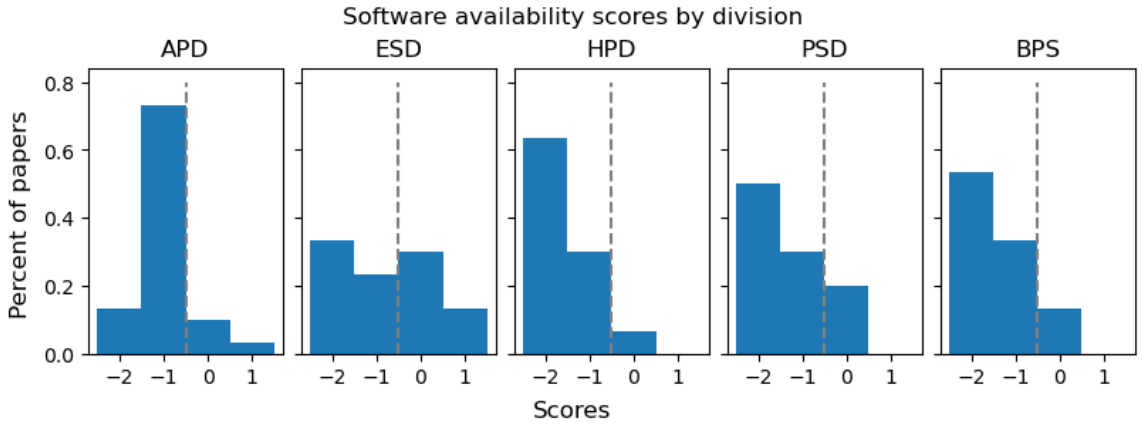

Figure 3: Software availability scores by division. Dotted lines signify the minimum requirements to meet compliance, with papers describing noncompliant practices to the left of the line and papers describing compliant practices to the right of the line.

DISCUSSION

Limitations

This study utilized publication data from 2023 as a baseline for assessing the implementation of SPD-41a requirements. However, it is important to acknowledge that articles published in 2023 are highly unlikely to reflect work covered by SPD-41a, which was implemented starting with the ROSES-2023 solicitation. The level of compliance with SPD-41a requirements demonstrated in 2023 publications likely reflect a combination of prior scientific information requirements incorporated into earlier ROSES solicitations and community-driven movement toward open science practices. The data underlying this report serve as a point of reference for future monitoring of SPD-41a implementation.

A significant limitation encountered in this study was the absence of a comprehensive list of all SMD-funded publications. This lack of a definitive catalog has several consequences. First, the data obtained from CHORUS and SciX do not allow for the precise differentiation between publications funded specifically by SMD versus those funded by NASA more broadly. Second, the inability to identify all SMD-funded publications precludes the calculation of precise compliance rates for public access to publications, as a reliable denominator for these calculations is unavailable.

A third limitation is that publications may be made openly available in manners that are not recorded or may be only made openly available at a later date. This may be via an agency repository, preprint service, on an author's website, or other means that are not recorded in the CHORUS database. As the initial analysis was done in June 2024, some of these publications may still be subject to a one-year embargo and might only be openly available after the period of initial study. As such, the verified open access estimate is likely a lower limit. Some of these methods may not meet current NASA requirements for being available in a NASA designated repository.

Factors associated with compliant data and software sharing practices

Some journals included in the data and software metrics require statements on open data and software availability. Observationally, these journals tended to have higher rates of compliance with SPD-41a requirements for data and software availability. This was particularly evident in Earth Science journals, where such requirements are more common. This observation suggests a potential opportunity to increase compliance across all SMD-funded research areas by encouraging the adoption of similar journal requirements. Promoting the inclusion of open data and software statements in publication guidelines could enhance adherence to SPD-41a principles by making data and software sharing an explicit part of the publication process and providing authors with a field to describe their open science practices.

Future work

This study establishes a critical baseline for monitoring the implementation of SPD-41a. Building upon this foundation, future years of measurement will be essential to track progress and assess the long-term impact of the policy. Longitudinal data collection will provide insights into trends, identify areas of improvement, and allow for the evaluation of interventions aimed at enhancing compliance.

Future research should also incorporate the evaluation of Open Science and Data Management Plans (OSDMPs). Analyzing these plans would provide valuable insights into researchers' intended practices, data management strategies, and software sharing plans. This information would aid in identifying potential barriers to compliance and allow for targeted support and resources to be provided to researchers.

The current study focused primarily on R&A activities. To gain a more holistic understanding of open science adoption within SMD, future metrics should be developed to evaluate other SMD-funded scientific activities, including missions. Expanding the scope of evaluation to include missions will provide a more comprehensive assessment of SPD-41a implementation across the entire directorate.

AUTHOR STATEMENT

This work was conducted by the SMD Office of the Chief Science Data Officer (OCSDO). Rebecca Michelson (NASA internship) led study design, data collection, and analysis. Holly Norton (Booz Allen Hamilton) contributed to the study report. Rachel E. Paseka (NASA) and Steven M. Crawford (NASA) contributed to the design, study, and report.

SUPPORTING MATERIALS

The data and Python scripts underlying this study are archived on Zenodo. The notebooks underlying this work are available on GitHub.

This research has made use of the Science Explorer, funded by NASA under Cooperative Agreement 80NSSC21M00561.

Data analysis was conducted using Python, leveraging the pandas library (version 2.2.3; McKinney, 2010) in Jupyter Notebook (version 7.2.2; Kluyver et al., 2016) and matplotlib (version Hunter 2007) for generating visualizations.

REFERENCES

CHORUS. (2024). Clearinghouse for the Open Research of the United States. Retrieved from https://www.chorusaccess.org

J. D. Hunter, "Matplotlib: A 2D Graphics Environment", Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007, https://doi.org/10.1109/MCSE.2007.55

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., & Willing, C. (2016). Jupyter Notebooks—a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), Positioning and Power in Academic Publishing: Players, Agents and Agendas (pp. 87-90). IOS Press. https://doi.org/10.3233/978-1-61499-649-1-87

McKinney, W. (2010). Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference (pp. 51-56).

Michelson, R, Paseka, R, Crawford, S. 2024. Monitoring Implementation of the NASA Science Mission Directorate Scientific Information Policy for NASA's Research Opportunities for Earth and Space Science. OSF Registries. https://doi.org/10.17605/OSF.IO/43U8W

NASA Science Mission Directorate. (2021). Scientific Information Policy for the Science Mission Directorate (SPD-41). https://smd-cms.nasa.gov/wp-content/uploads/2023/05/ScientificInformationpolicySPD_41.pdf

NASA Science Mission Directorate. (2022). Scientific Information Policy for the Science Mission Directorate (SPD-41a). https://smd-cms.nasa.gov/wp-content/uploads/2023/08/smd-information-policy-spd-41a.pdf

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. CreateSpace.

Wikipedia contributors, "Population proportion," *Wikipedia, The Free Encyclopedia,* https://en.wikipedia.org/w/index.php?title=Population_proportion&oldid=1292089454 (accessed May 8, 2025).

APPENDIX I: GLOSSARY

This report uses the following definitions, consistent with SPD-41a:

- Publications include documents released through print, electronic, or alternative media. This includes peer reviewed manuscripts, technical reports, conference materials, and books. This does not include laboratory notebooks, preliminary analyses, drafts of scientific papers or preprints, plans for future research, peer review reports, or communications with colleagues

- Scientific software is software, including computer programs in source and object code, that provides users some degree of scientific utility or produces a scientific result or service. This includes codes used for analysis, simulation or modeling software, libraries, notebooks, or other packages needed to reproduce scientific results. It also includes scripts developed to produce the data products, figures, and tables included in a scientific publication.

- Data are scientifically or technically relevant information that can be stored digitally and accessed electronically. This includes any scientifically useful data associated with an award, including the information needed to validate the scientific conclusions of publications. Data does not include laboratory notebooks, preliminary analyses, peer review reports, or communications with colleagues.

Open Science and Data Management Plan (OSDMP): A document that describes how scientific information produced from a scientific activity will be shared and preserved in accordance with relevant policies. The OSDMP should include plans for sharing data, software, and publications. It may also include a description of other types of scientific information that will be shared openly and

other open science activities associated with the project. An OSDMP replaces the data management plan (DMP) in some funding proposals.

APPENDIX II: RUBRICS FOR DATA AND SOFTWARE AVAILABILITY SCORES

Rubric for Data Availability

Evaluation	Score	Criteria
Exceeds Requirements	1	Data sharing is assessed to be compliant and data are made more accessible or reproducible. This may include being compliant with FAIR principles, integrated into larger data sets, or other practices.
Compliant	0	Data are shared to an appropriate repository and made publicly available. Data are made available in machine readable, non-proprietary, modifiable formats. Data are made available with a license and are citable with a persistent identifier. OR Sufficient description of why data produced by the study are not appropriate for public release.
Non-Compliant	-1	“Data will be shared upon request” OR Data are shared to a personal website, or to a repository that does not meet the guidelines for acceptable data repositories in Appendix D of SPD-41a OR

		Data are made publicly available via an appropriate repository but are not machine readable, are shared in a proprietary format, lack a license, or lack a persistent identifier.
None	-2	Data produced for the study are not made available
Not applicable	0	No data are produced. This includes the reuse of publicly available data.

Rubric for Software Availability

Evaluation	Score	Criteria
Exceeds Requirements	1	Software is assessed to be compliant and is made more accessible or reproducible. This may include documentation, testing, or other features that enhance reproducibility. This can also include contributing to open source projects. OR If no software is being developed, this score is not applicable.
Compliant	0	Software developed for this project is made openly available, citable with a persistent identifier, and archived. It includes the parameters necessary to replicate the process. It may be lacking a permissive license, written in a proprietary language, or include unreleased restricted software. OR If existing software was used, it was described well and necessary parameters to replicate the process were included.
Non-Compliant	-1	Software developed for this project is made openly available but includes some weaknesses (e.g., software not

		appropriately archived or citable with a persistent identifier). It may be missing parameters necessary to replicate the process. OR If existing software was used, it was described well but necessary parameters to replicate the process were not included.
None	-2	Software developed for this project is not made available, and parameters are not described. OR If existing software was used, it was not described.
Not applicable	0	No software was produced as part of the work.