

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

METRICS REPORT

Metrics on 2024 Publications, Data, and Software
associated with NASA Science Mission Directorate
Funding

Office of the Chief Science Data Officer
NASA Science Mission Directorate
February 17, 2026



TABLE OF CONTENTS

EXECUTIVE SUMMARY	1
INTRODUCTION.....	2
EVALUATION OF PUBLICATIONS	3
Publication Availability.....	3
EVALUATION OF SCIENTIFIC DATA AND SOFTWARE UNDERLYING PUBLICATIONS.....	5
Data and Software Availability.....	6
DISCUSSION	9
Limitations	9
Factors associated with compliant data and software sharing practices.....	9
Future work.....	10
AUTHOR STATEMENT	10
SUPPORTING MATERIALS.....	10
REFERENCES.....	11
APPENDIX I: GLOSSARY.....	12
APPENDIX II: RUBRICS FOR DATA AND SOFTWARE.....	13
Rubric for Data Availability	13
Rubric for Software Availability.....	14
APPENDIX III: IDENTIFICATION OF DIVISIONS FOR PUBLICATIONS	15



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 2024 with respect to the SPD-41a requirements. While most publications will be from solicitations from earlier than 2023, these metrics help establish a baseline for compliance with the SPD-41a requirements.

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 51% to 85%, reflecting differences in detection of NASA-funded publications and varying methods of making publications openly available. This shows slight increases from the metrics measured in 2023.
2. Over 2023-2024, statistical analysis of a representative sample of peer-reviewed publications across each division of SMD indicated that data availability ranged from 15% to 56% and software availability ranged from 13% to 33%.

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.

This report is an update of the [SPD-41a 2023 Publications report](#) and uses the same text and methodology throughout the report. Plans for this study are described in a preregistered report that was completed prior to data collection ([Michelson et al., 2024](#)). This study spanned Research and Analysis 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 [NASA CHORUS Dashboard](#). The dashboard provides analytical capabilities for NASA as a participating funding agency. 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 through a search of publications that acknowledge support of NASA. Publication metadata are aggregated by SciX from multiple sources, and while comprehensive, the publication record may be incomplete. This could be due to a lack of easily discernable acknowledgement to NASA or due to publications not yet being incorporated into SciX.

The analysis is performed on NASA-funded publications rather than only on SMD-funded publications due to the available information. Future analysis may focus on SMD-funded publications.

Publication Availability

The CHORUS database and NASA Science Explorer (SciX) were queried for peer reviewed publications that were supported by NASA awards. This report uses results as query in November 2025.

The [CHORUS database](#) reported 5,705 NASA-funded papers published in 2024. Of the 5,705 publications, 2887 (50.6%) were verified as open access

(OA; Figure 1) as reported by the publisher. As calculated from data from November 2025, these numbers are very similar to the statistics for 2023 with 5,798 publications and 51.8% being available as open access.

The SciX database reported a [total number of NASA-support peer-reviewed publications](#) as 5,019 in 2024. For those in SciX, 4,259 (85%) of publications are reported [as open access](#). This would include publications made openly available by the publisher and as preprints. These are similar numbers to 2023 where a [total number of NASA peer-reviewed publications](#) was 4642 with 86% being reported by the [NASA Science Explorer \(SciX\) database as open access](#).

These figures serve 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.

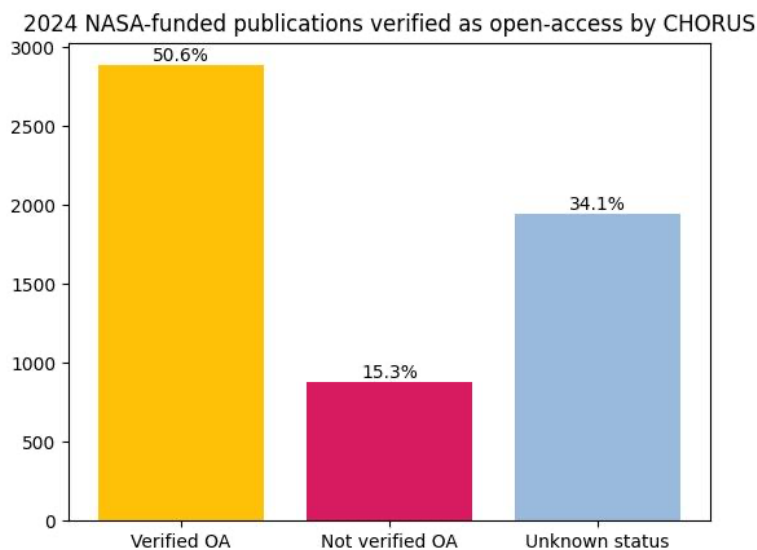


Figure 1: Percent of NASA-funded journal articles verified as open access (OA), where blue is unknown, gold represents the percentage of articles verified as OA, and crimson represents articles not verified OA.

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 2024.

From the CHORUS publication list, a random sample was generated to identify 150 SMD-funded publications, with 30 publications from each of the five SMD science divisions. Divisions were initially identified through classification via a machine learning algorithm trained on a set of publications labeled by the funding division (see Appendix III for more details). 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.

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.

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)	
	2023	2024	2023	2024
APD	20	15	13.3	16.6
BPS	40	36.7	13.3	18.3
ESD	80	56.7	40.3	33.3
HPD	20	26.6	6.7	13.3
PSD	36.7	35	20	18.3

Table 1: Percentage of sampled publications with underlying data and software shared in compliance with SPD-41a requirements for 2023 and 2024.

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. Figure 4 and Figure 5 show the compliance per division over the two years of analysis for data and software. This can serve as a baseline for future assessments.

All divisions showed publications that were compliant with the requirements in SPD-41a. with most of the publications in 2024 showed partial compliance with either the data or software sharing requirements. These findings suggest that progress continues to be 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.

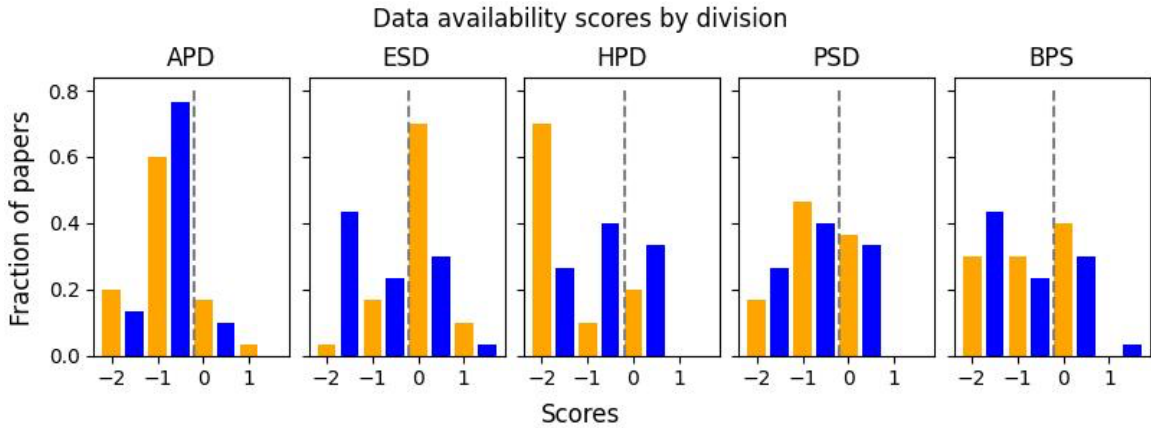


Figure 2: Data availability scores by division. Dotted lines signify the minimum requirements to meet compliance, with papers describing noncompliant practices to the left of the line and papers describing compliant practices to the right of the line. The orange bars are for 2023 publications and the blue bars are for 2024 publications.

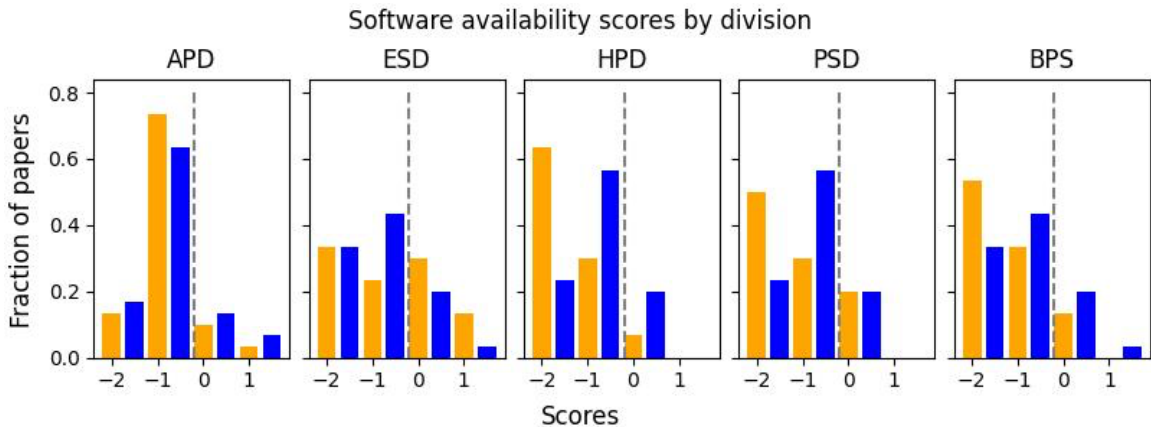


Figure 3: Software availability scores by division. Dotted lines signify the minimum requirements to meet compliance, with papers describing noncompliant practices to the left of the line and papers describing compliant practices to the right of the line. The orange bars are for 2023 publications and the blue bars are for 2024 publications.

Data Compliance Status by Division

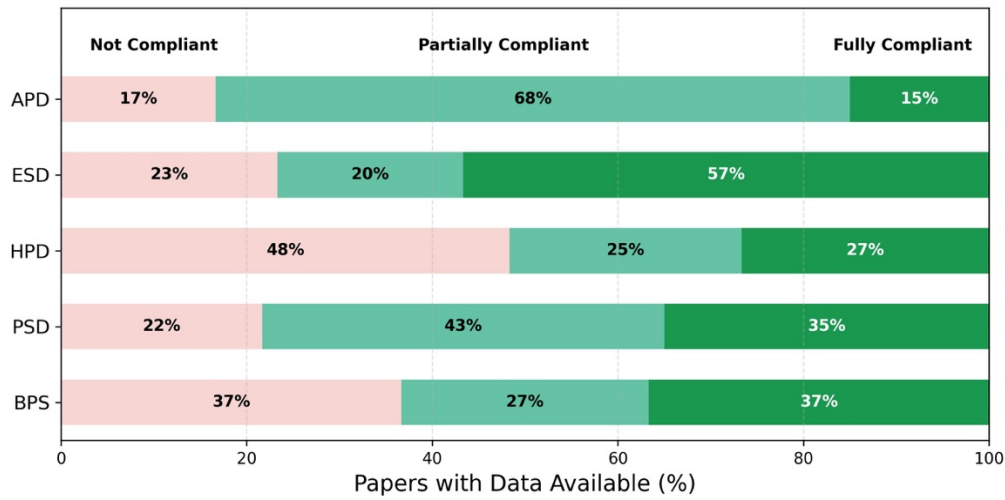


Figure 4: Data compliance status by division summed over the data from 2023-2024. Each of the divisions show a majority of papers are fully or partially compliant.

Software Compliance Status by Division

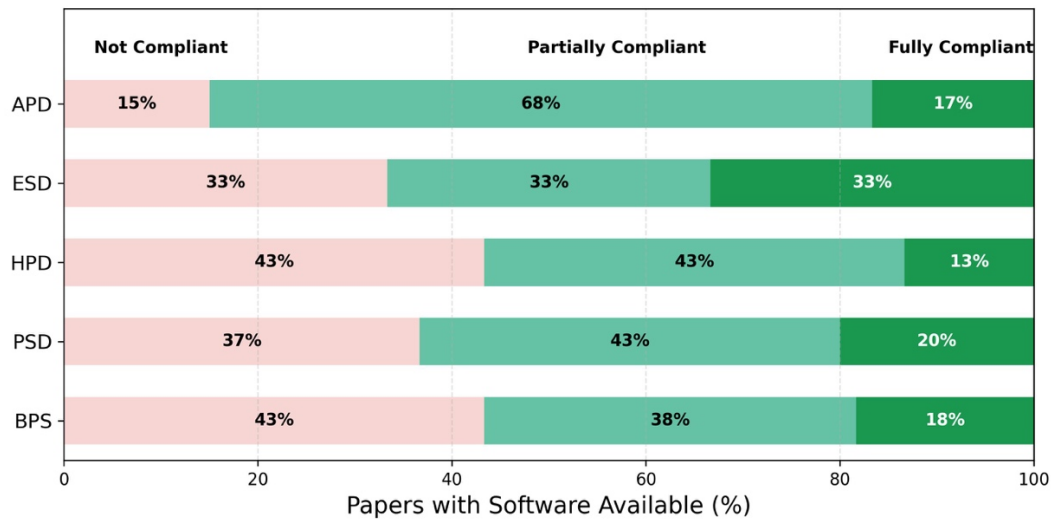


Figure 5: Software compliance status by division summed over the data from 2023-2024. Each of the divisions show a majority of papers are fully or partially compliant.

DISCUSSION

Limitations

This study utilized publication data from 2023 and 2024 as a baseline for assessing the implementation of SPD-41a requirements. The following limitations may restrict the precision of the results:

- SPD-41a requirements were only fully implemented in ROSES-2023 solicitation and almost all of the publications in this study are from programs supported under earlier solicitations
- The data from CHORUS and SciX may be incomplete or include publications not funded by SMD.
- Publications, data, or software may become openly available at different periods or the publications may not well describe the availability of data or software.

Due these limitations, the results reported here should serve as a point of reference for future monitoring of SPD-41a implementation and a baseline for future studies.

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 continues to establish 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.

AUTHOR STATEMENT

This work was conducted by the SMD Office of the Chief Science Data Officer (OCSDO) by Steven M. Crawford (NASA) with contributions from Chelle Gentemann (NASA/IPA).

This work acknowledges contributions from Rebecca Michelson (NASA internship), Holly Norton (Booz Allen Hamilton), Jeamay Palo (Science Systems and Applications, INC), and Rachel E. Paseka (NASA) to the [SPD-41a 2023 Publications 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 numpy (version 2.1.1; Harris et al. 2020) and 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

- Crawford, S., Paseka, R., Michelson, R. L., & Norton, H. (2025). Monitoring Implementation of the Scientific Information Policy for the NASA Science Mission Directorate (SPD-41a) Metrics Report. Zenodo.
<https://doi.org/10.5281/zenodo.15634481>
- CHORUS. (2024). Clearinghouse for the Open Research of the United States. Retrieved from <https://www.chorusaccess.org>
- Harris, C.R., Millman, K.J., van der Walt, S.J. et al., Array programming with NumPy, Nature, 585, 357–362, 2020, <https://doi.org/10.1038/s41586-020-2649-2>
- 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.

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.

APPENDIX III: IDENTIFICATION OF DIVISIONS FOR PUBLICATIONS

Divisions for each publication were identified using a machine learning model that was based on a data set that had identified the original funding source for a set of representative publications.

The data set matched publications from 2024 to the funding division through an internal data set of grant identification numbers and division funding. This included 1986 publications that were successfully matched to a single division. The data set was split into an independent training (80%) and testing (20%) data sets.

A simple two parameter model was then built to predict the division based on the input title and journal name of the publication. The journal titles were each treated as a unique entry, and a dictionary was built with the number of divisions funding a publication appearing in that journal. Titles were split into individual,

lower case words with non-alphabet characters removed. For each word, a dictionary was built with the number of divisions funding a publication with a title that uses that word. These dictionaries were then converted into distributions by normalizing by the total number of publications using that journal or word.

Based on the journal name and title, the model would return a probability for the publication to be funded by each division. This would be calculated from multiplying the probability calculated from the journal title and the probability calculated from the title. After different attempts to optimize the weighting of the probability from the two different measurements, it was found that the best results were found by setting the probably to one when the journal title was only associated with a single division.

With this model, the division of the publication was accurately predicted for 75% of the training data set. As divisions do have overlapping subject matter such as atmospheres and exoplanets, it is likely that this is close to the maximum accuracy that could be expected, but further study would be needed. This is likely sufficient classification for the purposes needed in this study that can be further verified by the reviewer.

The software and model are made available as [part of this release](#).