

Open-Source Science Guidance NASA Science Mission Directorate

December 8, 2022

2022 EXPLORE SCIENCE

Open-Source Science Guidance

NASA Science Mission Directorate December 8, 2022

Version History

#	Date	Description
1	12/08/2022	SMD produced the initial
		version of this guidance to
		provide additional information
		in support of the release of
		SPD-41a. This document will
		be updated as new
		information becomes
		available.

Background and Motivation	4
Limitations	4
Contributions	5
Open Science and Data Management Plan	6
Data Management Plan	6
Software Management Plan	6
Publication Sharing Plan	7
Other Open Science Activities	7
Sharing Publications	8
Requirements for Public Access to Publications	8
How to Share Publications	8
Open Access Publishing	9
Preprints	10
Data Management and Sharing	11
Data in Scope of SPD-41a	11
Data Management Plans	11
Timeline for Sharing Data	12
Where to Share Data	13
FAIR Data	14
Sharing Data	14
Data Policies and Guidance for SMD Divisions	15
Software Management and Sharing	16
Software in Scope of SPD-41a	16
Software Management Plans	17

Timeline for Sharing Software	18
Where to Share Software	18
General Expectations for Open-Source Software	19
Reporting Open-Source Software	20
Intellectual Property and Licenses for Software	20
Types of Research Software and Expectations for Sharing	20
Software Policies and Guidance for SMD Divisions	22
Sharing Materials from Science Events	23
Requirements	23
How to Share Publications Produced for an Event	23
Glossary of Open-Source Science Terms	24

Background and Motivation

NASA's Open-Source Science Initiative (OSSI) is a comprehensive program of activities to enable and support moving science towards openness, and it aims to implement SMD's Strategy for Data Management and Computing for Groundbreaking Science 2019-2024. OSSI includes a commitment to the open sharing of software, data, and knowledge (algorithms, papers, documents, ancillary information) as early as possible in the scientific process. The principles of open-source science are to make publicly funded scientific research transparent, inclusive, accessible, and reproducible.

As part of the OSSI, <u>SMD's Scientific Information Policy</u> provides guidance on the open sharing of publications, data, and software created in the pursuit of scientific knowledge. <u>SPD-41a:</u>

<u>Scientific Information Policy for the Science Mission Directorate</u> updates the previously released <u>SPD-41</u>, which consolidated existing Federal and NASA policy on sharing scientific information. SPD-41a was developed in light of new federal guidance on <u>Ensuring Free, Immediate, and Equitable Access to Federally Funded Research</u>, studies from the National Academies, and input from the SMD scientific community.

This document provides guidelines, best practices, and examples of open-source science to support the SMD scientific community in implementing the requirements of SPD-41a and achieving the broad OSSI goal of moving science towards openness. This guidance represents general recommendations and best practices for the SMD scientific community and should be considered alongside any additional guidance provided by SMD Divisions or specific funding solicitations. For additional information about SPD-41a and other scientific information policies that correspond to SMD Divisions, see the Scientific Information Policy website and Frequently Asked Questions (FAQ). For open science training opportunities and additional resources, see NASA's Transform to OPen Science (TOPS) program.

Limitations

This document does not establish policy requirements related to SMD funded activities. While the guidance described here is compliant with Science Mission Directorate, it may not be comprehensive or complete, and there may be other ways in which to comply with existing or future agreements or requirements not described here.

For requirements related to scientific information on SMD-funded activities, please see <u>SPD-41a</u>, NASA policies, and/or any requirements in the solicitation or agreement associated with the activity.

Contributions

This document will be updated as new information becomes available. Following the best practices of open science, this material can be best built through the collective knowledge of the community. To report something that needs to be fixed or to suggest an addition, please send an email to HQ-SMD-SPD41@mail.nasa.gov.

Open Science and Data Management Plan

An Open Science and Data Management Plan (OSDMP) describes how the scientific information that will be produced from SMD-funded scientific activities will be managed and made openly available. At a minimum, the OSDMP should include sections on data
management, software management, and publication sharing that describe how these categories of scientific information will be managed and openly shared to comply with applicable SMD scientific information policies. The OSDMP 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.

SPD-41a requires a data management plan and a software management plan for all SMD-funded scientific activities but does not require an OSDMP. However, many SMD funding opportunities starting with ROSES-2023 will require an OSDMP instead of the previous requirement for a data management plan.

OSDMPs should reflect the practices of specific research communities, and SMD Divisions and/or specific ROSES program elements may provide additional guidance on components such as data management plans and software management plans.

Data Management Plan

All SMD-funded scientific activities that are expected to produce scientific data shall include a data management plan describing how data will be managed, preserved, and released to comply with the requirements of SPD-41a. General guidance on the components of data management plans for SMD is provided in the Data Management and Sharing section of this guide.

Software Management Plan

All SMD-funded scientific activities that are expected to produce software shall include a software management plan describing how software will be managed, preserved, and released to comply with the requirements of SPD-41a. General guidance on the components of software management plans for SMD is provided in the Software Management and Sharing section of this guide.

Publication Sharing Plan

This section should describe the types of publications that are expected to be produced from the SMD-funded scientific activities, including peer reviewed manuscripts, technical reports, conference materials, and books. The plan should also outline the methods expected to be used to make the publications publicly accessible, which will likely include options listed under 'How to Share Publications' in this Open-Source Science Guidance.

Other Open Science Activities

The OSDMP may also include a description of additional open science activities associated with the project. This may include:

- Holding scientific workshops and meetings openly to enable broad participation
- Preregistering research plans in advance of conducting scientific activities
- Providing project personnel with open science training or enablement (if not described elsewhere in a proposal)
- Implementing practices that support the inclusion of broad, diverse communities in the scientific process as close to the start of research activities as possible (if not described elsewhere in a proposal)
- Integrating open science practices into citizen science activities
- Contributions to or involvement in open-science communities

Sharing Publications

The preservation of and public access to publications produced from SMD-funded scientific activities supports the scientific process, assures the widest dissemination of the information that NASA produces, and supports anyone, anywhere in reading the results of NASA scientific activities. This section contains supplemental guidance to support the implementation of requirements for public access to publications established by SPD-41a.

Requirements for Public Access to Publications

SPD-41a defines publications as scientific and technical 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.

Under SPD-41a, all publications resulting from research funded by SMD shall be made publicly accessible to the extent allowed by applicable law and existing NASA policies. Peer reviewed manuscripts resulting from SMD-funded scientific activities shall be made freely available to the public by default, without any embargo or delay after the publication date.

Authors may meet this requirement in a number of ways (see How to Share Publications), whether or not their manuscript is published as Open Access (see Open Access Publishing). Ultimately, SMD-funded publications must be made publicly available at the time of publication via the NASA Scientific and Technical Information (STI) Repository.

How to Share Publications

Authors have several options for how to make their publications publicly accessible in compliance with SPD-41a:

For articles that are published as Open Access (see Open Access Publishing), the final published article (*i.e.*, the publisher's version of record) can be made publicly available in the STI Repository.

For articles published as Open Access by journal publishers participating in the
 <u>Clearinghouse for the Open Research of the United States</u> (CHORUS), the published
 article will be made publicly available in the STI Repository on behalf of the authors.

Authors should verify that their article is available in the STI Repository following its publication, in which case no further action is required by the author. View a list of journal publishers participating in CHORUS.

- For articles published as Open Access that are indexed in the <u>NASA Astrophysics Data</u>
 <u>System</u> (ADS), no further action is required by the researcher to comply with public
 access requirements for the article at this time.
- For articles published as Open Access that are not covered by CHORUS or ADS, authors can submit the final published article to the NASA STI Repository via the PubSpace submission page at the time of the article's acceptance.

For articles that are not published as Open Access, the author's copy of an accepted manuscript can be made publicly available in the STI Repository. The accepted manuscript is the final, peer-reviewed version of the article that has been accepted for publication by a publisher. The accepted manuscript includes all changes made during the peer review process and contains the same content as the final published article, but it does not include the publisher's copyediting, stylistic, or formatting edits that will appear in the final journal publication (*i.e.*, the version of record). Authors can submit accepted manuscripts to the NASA STI Repository via the <u>PubSpace submission page</u> at the time of the article's acceptance.

Further examples for how to make publications available will be provided as they are developed.

Open Access Publishing

Open Access publications are defined here as those in which the publisher makes the article and related content available online, free of charges or other barriers to access, immediately at the time of publication. Such articles are often licensed for sharing and reuse under a <u>Creative Commons</u> or similar license-type agreement that allows the authors to retain copyright while allowing others to copy, distribute, and reuse their work. A short description of common Open Access publishing methods (*e.g.*, Gold or Diamond/Platinum) is available from <u>Framework for Open and Reproducible Research Training</u>.

Open Access publishing is not required to meet the public access requirements for peer-reviewed publications under SPD-41a. However, many publishers in Earth and Space science are moving toward Open Access models, and SMD-funded researchers are encouraged to publish their peer-reviewed manuscripts as Open Access (e.g., Gold or Diamond Open Access)

in reputable journals. Costs for Open Access publishing may be included in the budgets of proposals for SMD funding.

Preprints

SMD-funded researchers are encouraged to post their manuscripts to community-appropriate preprint servers such as <u>arXiv</u> or the <u>Earth and Space Science Open Archive</u>, though this is not a requirement under SPD-41a and does not currently satisfy requirements for public access as described above.

Data Management and Sharing

Consistent with the principles of open-source science, data resulting from SMD-funding shall be made publicly available through methods that are transparent, inclusive, accessible, and reproducible. This section contains supplemental guidance to support the implementation of requirements for data management and sharing established by SPD-41a. In addition to these pages, researchers should reference the guidance provided in the <u>data policies and guidance</u> provided by their relevant SMD Divisions. These policies provide specific information about where, how, and what data should be shared including specific data repositories, metadata requirements, and recommended data formats.

Data in Scope of SPD-41a

Data are defined as scientific or technically relevant information that can be stored digitally and accessed electronically. This includes any scientifically useful data associated with an award. In particular, the information needed to validate the scientific conclusions of peer-reviewed publications must be shared at the time of publication. This does not include laboratory notebooks, preliminary analyses, intermediate data products, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as laboratory specimens.

Data subject to specific laws, regulations, or policies (*e.g.*, Export Administration Regulations (EAR) or International Traffic in Arms Regulations (ITAR)) that would prevent the release of this information are exempt from requirements for making data publicly available. Section II-C of SPD-41a lists additional laws, regulations, and policies that generate exceptions to data sharing requirements.

Data Management Plans

All SMD-funded scientific activities that are expected to produce scientific data shall include data management plan describing the management, preservation, and release of data to facilitate implementation of relevant scientific information policies. The data management plan may be one component of a broader OSDMP. Though not required under SPD-41A, the OSDMP will be a required component of many proposals for SMD funding starting with ROSES-2023.

At a minimum, a data management plan for SMD-funded research should include:

- Descriptions of the data expected to be produced from the proposed activities, including types of data to be produced, the approximate amount of each data type expected, the machine-readable format of the data, data file format, and any applicable standards for the data or associated metadata
- The repository (or repositories) that will be used to archive data and metadata arising from the activities and the schedule for making data publicly available
- Description of data types that are subject to relevant laws, regulations, or policies that exclude them from data sharing requirements
- Roles and responsibilities of project personnel who will ensure implementation of the data management plan

Data management plans should reflect the practices of specific research communities, and SMD Divisions and/or ROSES program elements may provide additional guidance on format and content. Specific guidance includes:

- Planetary Science Division guidance and template
- Heliophysics Division guidance and template
- Earth Science Division guidance

Researchers may benefit from using online, open-source tools such as <u>DMPTool</u> to create their data management plans.

Timeline for Sharing Data

Scientific data resulting from SMD-funded scientific activities shall be made publicly available, to the extent allowed by applicable law and existing NASA policies, according to the following timeline:

 Scientific data needed to validate the scientific conclusions of peer-reviewed manuscripts resulting from SMD-funded scientific activities shall become publicly available no later than the publication date of the corresponding peer-reviewed article.

- This includes data required to derive the findings communicated in figures, maps, and tables, as well as scientifically useful data from models and simulations.
- Any scientifically useful data associated with a SMD research award shall be made
 publicly available by the end of the period of performance, whether or not the data would
 be needed to validate the scientific conclusions of a peer-reviewed publication.

Where to Share Data

Data must be shared and archived in locations that ensure the accessibility and preservation of the data.

Researchers should follow the guidance for how to share data as described in the relevant solicitation or division data policies. For some divisions, there are SMD-designated, domain specific data repositories (e.g., GeneLab, NExScl Science Data Archives, HLSP Data Collections) in which data developed under those programs should be shared. NASA's Science Data page includes links with more information on each Division's data repositories. Depending on the repository, this may require the data to be uploaded directly into the repository or this may be uploaded as part of the publication process.

If there is no designated NASA repository, options for where to share research data can include:

- In a non-SMD, federally supported data repository (e.g., <u>data.nasa.gov</u>). If using a different federally supported repository, this should be indicated in the data management plan.
- In public repositories already in use by the scientific community that have characteristics
 consistent with <u>Desirable Characteristics of Data Repositories for Federally Funded</u>
 <u>Research</u>; further examples will be provided but see examples provided by the <u>NIH</u> and USGS.
- If appropriate for the field or journal, as machine readable tables in the supplemental
 material of a peer-reviewed publication. This may be the best solution for small data sets
 or individual tables that accompany a peer-reviewed publication, but the use of a
 community-recognized repository is encouraged.

The method for sharing the data must be described as part of the data management plan.

Especially for very large data sets that might not fit into existing guidance, the PI must describe

in their DMP how they will share them, what is appropriate to share, and include the costs for sharing them in their budget.

FAIR Data

When sharing data, researchers should follow the <u>FAIR Guiding Principles</u> for scientific data management and stewardship. The FAIR Principles include ensuring that data are:

- <u>Findable</u> consistent and persistent descriptions make scientific data easy to find by both humans and computers
- Accessible use of standard, open protocols ensure data and metadata can be accessed by all.
- Interoperable formal, accessible, and widely adopted semantics and vocabularies are used to expand data usability across systems and communities
- Reusable data are richly described according to standards to ensure they can be combined or replicated, and usage rights are clarified

The **Sharing Data** guidance below is consistent with the FAIR Guiding Principles.

Sharing Data

The following items are required (denoted using "shall") or recommended (denoted using "should") when publicly sharing research data developed using SMD funding. These items help ensure that data are preserved and accessible to support reproducibility and reuse, and they are consistent with the <u>FAIR Guiding Principles</u>.

Open accessibility: Publicly available, SMD-funded data shall be made available without fee or restriction of use. The data shall be shared in a repository that provides broad, equitable, and maximally open access to datasets and their metadata free of charge in a timely manner after submission, consistent with legal and policy requirements related to maintaining privacy and confidentiality, Tribal and national data sovereignty, and protection of sensitive data. The data will be accessible to the public (lay and scientific) without pre-approval.

- <u>Format:</u> SMD-funded data and metadata shall be made available for access, download, or export in non-proprietary, modifiable, open, and machine-readable formats consistent with standards used in the disciplines the repository serves.
- <u>Inclusion of metadata:</u> SMD-funded data shall include robust, standards-compliant metadata that clearly and explicitly describe the data. Inclusion of metadata by the repository enables discovery, reuse, and citation of SMD-funded data.
- Clear guidance on use: Publicly available SMD-funded data shall be reusable with a clear, open, and accessible data license. If there are no other restrictions on the data, the data should be released with a Creative Commons Zero License. This provides a clear license for the user that the scientific data is in the worldwide public domain and that they may use it freely. In some cases, there might be existing restrictions on releasing the data due to intellectual property rights, contract restrictions, underlying licenses, or other issues. If unsure, contact your counsel that can help with intellectual property rights or ask for clarifications at HQ-SMD-SPD41@mail.nasa.gov.
- <u>Persistent identifiers:</u> Publicly available SMD-funded data collections shall be citable using unique persistent identifiers (*e.g.*, DOI) assigned by the repository to support data discovery, reporting, and research assessment.
- <u>Findability:</u> SMD-funded data shall be findable, such that the data can be retrieved, downloaded, indexed, and searched. The data must be shared in a repository that will ensure that data are searchable and be provided with descriptive metadata along with the data collections.

Data Policies and Guidance for SMD Divisions

Biological and Physical Science Division

Earth Science Division

Heliophysics Division

Software Management and Sharing

Consistent with the principles of open-source science, software developed using SMD funding shall be made publicly available through methods that are transparent, inclusive, accessible, and reproducible. The open sharing of research software improves the reproducibility of research, enables other scientists to use and build upon software developed using public funds, and allows the researchers who share the software to be cited and recognized for the impact of their work. As software has become increasingly more important to the scientific process, scientific manuscripts do not always capture all of the nuance that is in the source code that supports the paper. Sharing the software ensures that others can better understand and reproduce the results reported in the manuscript.

This page contains supplemental guidance for researchers to support the implementation of requirements for software management and sharing established by SPD-41a. In addition to these pages, researchers should reference the guidance provided in the <u>software policies and guidance</u> provided by their relevant divisions.

Software in Scope of SPD-41a

Software is defined as computer programs, including source and object code, that provide users some degree of utility or service. Scientific software in scope of SPD-41a is software that provides users some degree of scientific utility or produces a scientific result or service. Further definitions and examples of software are provided in Types of Research Software and Expectations for Sharing. The guidance provided here is most relevant for research software that is produced by investigations funded via SMD research awards. This software should be developed and released as described in the project's software management plan.

Software developed only for preliminary analysis, plans for future research, or communication with colleagues is not required to be released.

Restricted software, that is software that is subject to specific laws, regulations, or policies (*e.g.*, Export Administration Regulations (EAR), International Traffic in Arms Regulations (ITAR), intellectual property laws, license restrictions) that would prevent the release of this information, is exempt from the requirements for making software publicly available. Section II-C of SPD-41a lists additional laws, regulations, and policies that generate exceptions to software sharing

requirements. This includes scientific software for which release is limited by patent rights, as described in the governing document of the funding mechanism, including "Patent Rights for Small Business Firms and NonProfit Organizations."

Commercial software, that is software that is produced for the purposes of sale and includes software that would be classified as commercial-off-the-shelf (CoTS), is not included in the types of software that must be released as part of research awards. Software developed in a proprietary or commercial language, such as IDL or MATLab, is expected to be released if allowed by the license.

Software Management Plans

All SMD-funded scientific activities that are expected to produce software shall include a software management plan (SMP) describing how software will be managed, preserved, and released to comply with the requirements of SPD-41a. The software management plan may be one component of a broader OSDMP. Though not required under SPD-41A, the OSDMP will be a required component of many proposals for SMD funding starting with ROSES-2023.

At a minimum, a software management plan for SMD-funded research should include:

- Descriptions of the software expected to be produced from the proposed activities, including types of software to be produced, how the software will be developed, and the addition of new features or updates to existing software. This can include the platforms used for development, project management, and community-based best practices to be included such as documentation, testing, dependencies, and versioning.
- The repository(ies) that will be used to archive software arising from the activities and the schedule for making the software publicly available
- Description of software that are subject to relevant laws, regulations, or policies that exclude them from software sharing requirements
- Roles and responsibilities of project personnel who will ensure implementation of the software management plan

Software management plans should reflect the practices of specific research communities, and SMD Divisions and/or ROSES program elements may provide additional guidance on format and content.

Timeline for Sharing Software

Scientific software resulting from SMD-funded scientific activities shall be made publicly available, to the extent allowed by applicable law and existing NASA policies, according to the following timeline:

- 1. Scientific software needed to validate the scientific conclusions of peer-reviewed manuscripts resulting from SMD-funded scientific activities shall become publicly available no later than the publication date of the corresponding peer-reviewed article. This includes software required to derive the findings communicated in figures, maps, and tables, as well as scientifically useful software from models and simulations.
- Any scientifically useful software associated with a SMD research award shall be made publicly available by the end of the period of performance, whether or not the software would be needed to validate the scientific conclusions of a peer-reviewed publication.

Where to Share Software

Software must be shared and archived in locations that ensure its accessibility and preservation. Researchers should follow the guidance for how to share software as described in the relevant solicitation or under the division software policies if available.

If there is no specific guidance on how to share software, options for where to share software can include:

- In the supplemental material of a publication. This is ideal for small scripts, notebooks, or spreadsheets that include calculations necessary for reproducing the paper.
- Publishing the source code in a software specific journal.
- If shared on a version controlled platform (e.g. <u>GitHub</u>), it is also important to archive the software at a designated repository.
 - GitHub has integration with Zenodo to make it easy to create an archived copy of the software.
- In public repositories already used in the community such as <u>Zenodo</u>, <u>Astrophysics</u>
 <u>Source Code Library</u>, and <u>Software Heritage</u>.

The method for archiving and sharing the software should be described in the Software Management Plan.

General Expectations for Open-Source Software

Unrestricted software developed using SMD-funding is expected to be shared openly. There are many different types of software and the expectations for software sharing are different for missions and researchers (See Types of Research Software and Expectations for Sharing).

When released, SMD-funded software should follow best practices in the relevant open source and research communities. For example, providing documentation and testing, which are not required to be provided under SPD-41a, alongside the source code increases the quality of the software and reusability of the software.

For publicly available software projects, SMD-funded software projects must include a code of conduct and guidelines for how to make contributions. A code of conduct provides clear guidelines for the conduct of those participating in the development of the software. The guidelines for how to make contributions provide clear expectations for how to contribute to the project. This may include how to make contributions, the type of contributions that the project is accepting, or even that the project is currently not accepting contributions. This can also include the expectations for support for the software project and for responding to questions about the project.

When released as open source software, source code for SMD-funded software shall be made available in a publicly accessible repository that is widely recognized by the community. See the section on 'Where to Share Software' for more information and examples of how to make the software available.

Publicly available SMD-funded software must be citable using a persistent identifier. A persistent identifier such as a DOI provides an easy method for software to be cited in the scientific literature and for the developers of the software to receive credit for their work. It also provides a way to track the usage of the software and to make reporting on the software easy.

For software developed under research grants, there is no expectation that the software is maintained. Some scientific software is developed for a single purpose, and there is no benefit in further development or maintenance of the software. Providing it openly does help to further the understandability of the scientific work that it supports as a manuscript may not capture all of the details in the processing or analysis required to reproduce the results.

Reporting Open-Source Software

Publicly available SMD-funded software shall be indexed as part of the NASA catalog of software. Developers of software packages that are developed as part of SMD-funded activities must catalog the software in in NASA's New Technology Reporting System. Single use software and commercial software do not need to be reported for indexing as part of the NASA catalog of software.

Intellectual Property and Licenses for Software

If there are no other restrictions, publicly available SMD-funded software should be released under a permissive license that has broad acceptance in the community. Restrictions that may prevent release under a permissive license include, but are not limited by, software governed by incompatible licenses or inclusion of restricted computer software. Seek specific advice from the Chief Science Data Office or Intellectual Property Counsel, as needed. Questions can be submitted to HQ-SMD-SPD41@list.nasa.gov.

For software developed at NASA Centers and released through the NPR 2210 process, Center Intellectual Property Counsel shall be consulted in the selection of the license to be used in the release of software, which may include Apache 2.0, BSD, or MIT.

Types of Research Software and Expectations for Sharing

Different types of research software have different expectations and practices for being shared openly, as defined in Appendix F of SPD-41a.

Single use: Software written for use in unique instances, such as making a plot for a paper or manipulating data in a specific way. Single-use software developed from SMD-funding shall be released no later than with the publication it supports. It can be released as supplementary material to the publication or shared in a data repository that produces a permanent identifier that is linked to from the publication.

Libraries: Generic tools, often with a larger user base, implementing well known algorithms, providing statistical analysis or visualization, etc., that are incorporated in other software categories. Libraries developed using SMD funding shall be released no later than the publication it supports. Libraries should be developed openly in a version-controlled platform and SMD-funded researchers are encouraged to contribute to existing, open source libraries. Major versions or releases of the software should be archived in a NASA designated repository.

Analysis software: Generalized software (not low-level libraries) used to manipulate measurements or model results to visualize or gain understanding. This software often evolves from single-use utility software and may incorporate libraries. Analysis software developed from SMD-funding shall be released no later than the publication it supports. Analysis software should be developed openly in a version-controlled platform and SMD-funded researchers are encouraged to contribute to existing, open source analysis software. Major versions or releases of the software should be archived in a NASA designated repository.

Model and Simulation software: Software that either implements solutions to mathematical equations given input data and boundary conditions or infers models from data. Includes first-principles models, data-assimilation tools, empirical models, machine learning, mission planning and engineering tools, among others. They often use libraries.

Model and Simulation software developed from SMD funding shall be released no later than the peer-review publication it supports. Model and Simulation software should be developed openly in a version controlled platform and SMD-funded researchers are encouraged to contributed to existing, open source software. Major versions or releases of the software should be archived in a NASA designated repository.

Data processing: Software for processing uncalibrated sensor measurements into calibrated sensor data and derived data products. This software type applies calibration coefficients, corrections or algorithms, which may be dependent on forward modeling, simulated observations, equations, and data filtering. It may include modeling and simulation software and libraries.

Data processing software developed under research grants from SMD funding shall be released no later than the peer-review publication it supports. Data processing software should be developed openly in a version-controlled platform and SMD-funded researchers are encouraged to contribute to existing, open source software. Major versions or releases of the software should be archived in a NASA designated repository.

Commercial software: Software produced for the purposes of sale. This includes software that would be classified as commercial-off-the-shelf (CoTS) and software that NASA does not have a license to distribute. Commercial software is not required to be released. SMD-funded research software developed in a commercial language (a programming language that requires a license to compile or run software) is not considered commercial software and shall be shared in a similar manner as other types of software developed in open languages.

Software Policies and Guidance for SMD Divisions

Earth Science Data System Open Source Software Policy

Sharing Materials from Science Events

The movement toward open-source science requires a culture shift to make science more inclusive, transparent, and collaborative at all stages of the scientific process, including scientific conferences, workshops, and symposia. This page contains supplemental guidance for researchers to support the implementation of requirements for sharing scientific publications produced for science events established by SPD-41a. In the future, this guidance will be expanded to include additional recommendations for event organizers on making SMD-funded science events more inclusive, accessible, and open.

Requirements

Under SPD-41a, participants sponsored with SMD funding to attend conferences, workshops, or symposia shall deposit their public scientific publications produced for the event in an appropriate NASA designated repository. This includes conference proceedings, slide decks, poster presentations, and any other publications produced for the event.

This requirement does not apply to restricted information, nor does it apply to laboratory notebooks, preliminary analyses, drafts of scientific papers or preprints, plans for future research, peer review reports, or communications with colleagues.

How to Share Publications Produced for an Event

A specific repository may be identified by a SMD solicitation. If no repository is specified, event participants have several options for sharing their scientific publications produced for SMD-funded events. These include:

- Submitting to <u>NASA STI Repository</u>. For those with a NASA identity who have completed the STRIVES process for their scientific publications produced for an event, these publications will be archived in the NTRS and no additional action is required by the researcher.
- Sharing through an external, community-accepted repository (e.g., Zenodo or arXiv).
 This option is recommended for those without a NASA identity.

Glossary of Open-Source Science Terms

Most terms in this section are copied from Appendix B of <u>SPD-41a</u>. Additional terms are noted here with an asterisk (*).

Archive: The process of storing data to ensure long term retention.

Accessible: As per the definition in the <u>FAIR</u> principles, data are retrievable by their identifier using standardized communications protocols.

Commercial Software: Software produced for the purposes of sale. This includes software that would be classified as commercial-off-the-shelf (CoTS) and software that NASA does not have a license to distribute.

Data: Scientific or technically relevant information that can be stored digitally and accessed electronically.

- Information produced by missions include observations, calibrations, coefficients, documentation, algorithms, and any ancillary information. Further definitions of mission data and examples are provided in Appendix E.
- Information needed to validate the scientific conclusions of peer-reviewed publications. This includes the data required to derive the findings communicated in figures, maps, and tables.
- This does not include laboratory notebooks, preliminary analyses, intermediate data products, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as laboratory specimens.

Data Management Plan (DMP): A document that describes whether and how data will be shared and preserved. A DMP should be compliant with NASA policies.

Findable: As per the definition in the <u>FAIR</u> principles, metadata and data should be easy to find for both humans and computers.

Interoperable: As per the definition in the <u>FAIR</u> principles, data are able to work with other applications or workflows for analysis, storage, and processing.

- * Machine readable: Refers to information or data that is in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost. (Source: https://resources.data.gov/glossary/machine-readable-file/)
- * Open Access publication: Although there are various uses of the term 'open access', this guidance document uses the term to refer to a method of publication in which the publisher makes the article and related content available online, free of charges or other barriers to access, immediately at the time of publication. Such articles are often licensed for sharing and reuse under a Creative Commons or similar license-type agreement that allows the authors to retain copyright while allowing others to copy, distribute, and reuse their work.

Open Format: A file format for storing digital data that is defined by an openly published specification, usually maintained by a standards organization and that can be used and implemented by anyone. Open file format is licensed with an open license.

Open Source Software (OSS): Software that can be accessed, used, modified, and shared by anyone. OSS is often distributed under licenses that comply with the definition of "Open Source" provided by the <u>Open Source Initiative</u> or meet the definition of "Free Software" provided by the <u>Free Software Foundation</u>.

Persistent Identifier: A long-lasting reference to a digital source. The digital object identifier (DOI) system is an example of a persistent identifier.

Permissive License: Guarantees the free use, modification, and redistribution of software, while still permitting proprietary derivative works. Examples include the <u>Apache License 2.0</u>, the <u>BSD 3-Clause "Revised" License</u>, and the <u>MIT License</u>.

Publication: Document 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.

Report: Document produced through print, electronic, or alternative media containing scientific and technical information. These documents are typically not peer reviewed. Examples include technical publications, technical memorandums, contractor reports, conference publications, special publications, and technical translations. It does not include interim research grant reports.

Repository: An organized storage location that makes data findable and accessible.

Restricted Software: Software that shall not be released due to an existing Federal law or guidance, NASA policy, or security concern. This includes software supporting security requirements described in STD-1006. For Mission software, projects should engage with the software release authority to determine status. Examples of software that may be restricted are command related software, instrument control, authentication, or communication software.

Review: Assessment of SMD-funded activities by an individual or group. This may include a panel, standing review board, or senior review.

Reusable: As per the definition in the <u>FAIR</u> principles, metadata and data should be well-described so they can be replicated and/or combined in different settings. This includes releasing the data with a clear and accessible data usage policy.

Scientific Information: Scientific knowledge produced as part of a research activity. This can include, but is not limited to, publications, data, and software.

Scientific Software: Software that provides users some degree of scientific utility or produces a scientific result or service.

Scientific Utility: Information that is necessary to validate research findings or beneficial for future research activities.

Single Use Software: Software written for use in unique instances, such as making a plot for a paper or manipulating data in a specific way.

Software: Computer programs in source and object code that provide users some degree of utility or produce a result or service.

Software Project: An activity to develop software. A software project typically has a version control platform on which development can occur collaboratively.

Source Code: A human-readable set of statements written in a programming language that together compose software. Programmers write software in source code, often saved as a text file on a computer. The terms code and source code are often used interchangeably.