

National Aeronautics and
Space Administration



EXPLORE SCIENCE

Science Information Policy

**Steve Crawford, Kevin Murphy,
Brad Carpenter, Patrick Koehn,
Patricia Knezek, and Megan
Ansdell**

26 January 2022

SMD Strategy for Data Management and Computing for Groundbreaking Science 2019-2024



Science Mission Directorate's
Strategy for Data Management and Computing for Groundbreaking Science 2019-2024

Prepared by the Strategic Data Management Working Group

Approved by:

A blue ink signature of Thomas H. Zurbuchen.

14/17/19

Thomas H. Zurbuchen, Ph.D.
Associate Administrator,
Science Mission Directorate

Vision: To enable **transformational open science** through the continuous evolution of science data and computing systems for NASA's Science Mission Directorate.

Mission:

- Lead an **innovative and sustainable program** supporting NASA's unique science missions with academic, international, and commercial partners to **enable groundbreaking discoveries with open science**.
- **Continually evolve systems** to ensure they are usable and support the latest analysis techniques while protecting scientific integrity.

Goal 1: Develop and Implement Capabilities to Enable Open Science

Goal 2: Continuous Evolution of Data and Computing Systems

Goal 3: Harness the Community and Strategic Partnerships for Innovation

SMD Strategy for Data Management and Computing for Groundbreaking Science 2019-2024

Goal 1: Develop and Implement Capabilities to Enable Open Science

1.1

Develop and implement a **consistent open data and software policy** tailored for SMD

1.2

Upgrade capabilities at existing archives to **support machine readable data access using open formats and data services**

1.3

Develop and implement a SMD data catalog to support discovery and access to complex scientific data across divisions

1.4

Increase transparency into how science data are being used through a free and open unified journal server

Goal 2: Continuous Evolution of Data and Computing Systems

2.1

Establish **standardized approaches for all new missions** and sponsored research that encourage the adoption of advanced techniques

2.2

Integrate investment decisions in High-End Computing with the strategic needs of the research communities

2.3

Invest in capabilities to use commercial cloud environments for open science

2.4

Invest in the tools and training necessary to enable breakthrough science through application of AI/ML

Goal 3: Harness the Community and Strategic Partnerships for Innovation

3.1

Develop **community of practice and standards group**

3.2

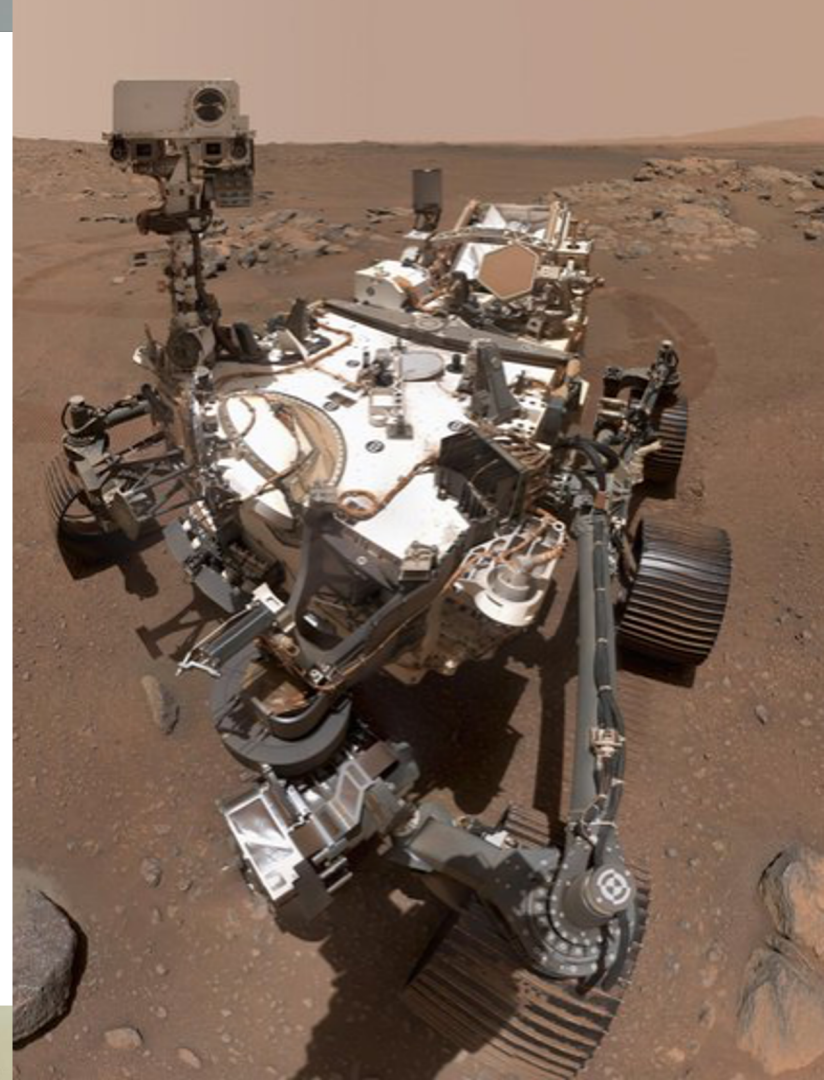
Partner with **academic, commercial, governmental and international organizations**

3.3

Promote opportunities for continuous learning as the field evolves through collaboration

Benefits of Open Science

- **Broadens participation** and **fosters greater collaboration** in scientific investigations by lowering barriers to entry
- Promotes **transparency** and **reproducibility**
- **Increases accessibility** to the data, software, and publications
- **Shortens the time** it takes for a new user to find and learn how to use data
- **Increases impact** including citations, media attention, and funding opportunities
- **Enables scaling** of large datasets for societal benefit and improves data systems efficiency





Core Values in the Development and Implementation of SPD-41

- Adhere to existing laws, policies and regulations to maximize the **openness** of useful **scientific information** produced by NASA SMD funding
- **Preserve** the scientific information produced by NASA SMD funding
- **Grow** the community that can access NASA's scientific information

Minimize burden in complying with the policy through supporting capabilities and staggered implementation

**As open as possible,
as closed as necessary**

Applicability of the Policy

All **future awards** will be in compliance SPD-41 - policy has been integrated into ROSES22 and SALMON calls

- PIs should include these costs in proposals

Current missions and grants encouraged to adopt SPD-41 ***to the extent possible*** within current resources or as directed by program officers ***for scientific useful information.***

SPD-41: Scientific Information Policy

SPD-41 is a consolidation of existing U.S. Government policies and laws applicable to SMD scientific information. These policies are based on existing NASA and Federal guidance, and in many instances are already part of solicitations for funding such as ROSES or SALMON Announcement of Opportunities.

This applies to all SMD-funded activities related to producing scientific information, but the policy excludes restricted information such as ITAR, export control, CUI.

- SPD-41: The Science Information Policy - <https://go.usa.gov/xtNTJ>
- Science Information Policy Website - <https://go.usa.gov/xtNTt>

SMD has released a Request for Information (RFI) on proposed additions to SPD-41 (SPD-41a) based on new Federal guidance, NASA policy, National Academy studies, or community best practices. **Responses are due March 4, 2022. <http://go.nasa.gov/RFISPD41>**

NASA Scientific Information

The information produced as part of NASA's scientific research activities represents a significant public investment. NASA holds this information as a public trust to increase knowledge and serve the public good.

The policy is focused on sharing **scientifically useful** information, such as:

- **Publications:** Scientific and technical documents released through print, electronic, or alternative media.
- **Data:** Scientific information that can be stored digitally and accessed electronically.
- **Software:** Computer programs in both source and object code that provide users some degree of scientific utility or produce a scientific result or service.

Publications are openly accessible

Making publications openly accessible can significantly lower the barriers to accessing tax-payer funded science.

- **An open access version of the as-accepted manuscript must be accessible via a NASA designated repository.**
- **We encourage publishing manuscripts as open access** and posting the as-accepted manuscript on community preprint services.
- In addition to peer-reviewed publications, **technical documentation should also be made publicly accessible.**
- **Open access publications have more citations, greater impact, and make our science more accessible.**

Data is publicly available without fee or restriction of use

Extending NASA's legacy of freely-available mission data to make our research data as accessible as possible.

- Research data underlying tables and figures becoming publicly available no later than the publication. This could be as part of supplemental material or in community appropriate repositories.
- Committed to the full and open sharing of scientifically useful information produced by SMD Missions with no period of exclusive access.
- Open access data increases impact, democratizes access, improves quality, and enables more science.

Software should be released as open-source software with a permissive license

Scientific discovery, understanding, validation, and interpretation are all enhanced by access to the source code of the software used by scientists.¹

- Research software developed using SMD funding and used in support of a scientific, peer-reviewed publication should be released as open source software no later than the publication date.
- Mission software related to scientific analysis will be made publicly accessible.
- Open source software enables reproducibility, improves understanding, and increases efficiency through the reuse of software.

New Guidance and Additional Policies (SPD-41a)

There are a range of new laws, recommendations, policies, and Federal Guidance related to Open Science that are applicable to the SMD community.

As such, **SPD-41a** incorporates this new guidance and makes them available for your comment prior to adoption.

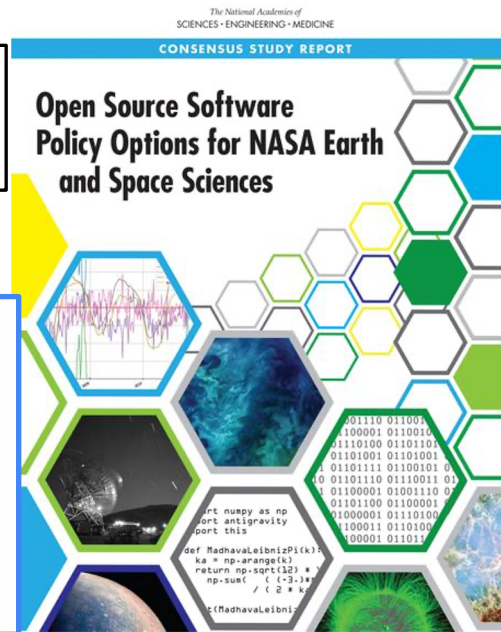
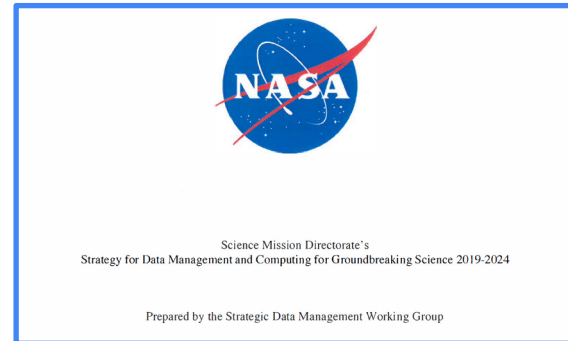
We need your help!

An RFI is open for comments on their impact and implementation:
<http://go.nasa.gov/RFISPD41>

TITLE II—OPEN GOVERNMENT DATA ACT

SEC. 201. SHORT TITLE.

This title may be cited as the “Open, Public, Electronic, and Necessary Government Data Act” or the “OPEN Government Data Act”.



Highlights of Proposed Additions

- SMD-funded data should follow the FAIR Guiding Principles for scientific data management and stewardship. This means data should be: **Findable, Accessible, Interoperable, and Reusable (FAIR)**.
- Persistent identifier should be used for funding mechanisms, missions, and data collections.
- SMD-funded investigators should have a persistent identifier, such as ORCID

Proposed additions are applicable to future solicitations, and awards and please respond RFI on their potential impact <http://go.nasa.gov/RFISPD41>

Highlights of Proposed Additions

- **Mission software shall be developed openly in a publicly accessible, version-controlled platform that allows for contributions and engagement from the community.**
- At the end of a research award, scientifically useful data associated with the award that has not already been made public shall be made publicly available.

Proposed additions are applicable to future solicitations, and awards and please respond RFI on their potential impact <http://go.nasa.gov/RFISPD41>

Highlights of Proposed Additions

- During SMD reviews, **peer reviewed data and software shall be recognized as having the commensurate value as peer reviewed manuscripts.**
- Conferences, workshops, and symposia for which SMD is the primary sponsor shall make **information produced during the conference publicly accessible.**

Proposed additions are applicable to future solicitations, and awards and please respond RFI on their potential impact <http://go.nasa.gov/RFISPD41>

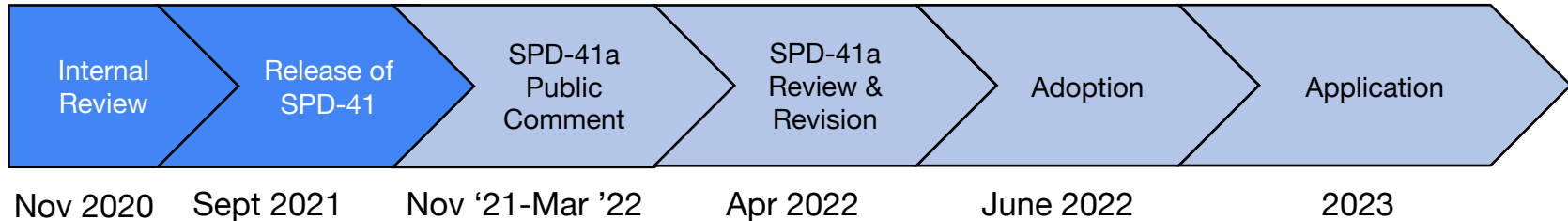
Applicability of the Policy

All **future awards** will be in compliance with existing policies that will be integrated into ROSES and SALMON calls.

- PIs should include these costs in proposals

Current missions and grants are encouraged to adhere to policies under their current contract or grant terms and should adopt ***to the extent possible*** new policies within current resources or as directed by program officers ***for scientific useful information***.

Schedule for policy development



SPD-41 was released in September 2021.

- Approval of SPD-41a will occur no earlier than June 2022.
- SPD-41a is expected to be in place for ROSES and AO calls in 2023.
- Some solicitations or divisions may adopt parts of the policy earlier.
- Existing Missions and Researchers should adopt the policy as resources allow.

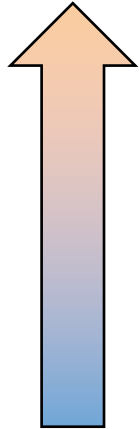
Researchers funded from ROSES23 will need to be compliant.

- Funded projects will typically start in 2024.
- Funded publications will typically be produced starting in 2025
- Compliance mechanisms to be investigated with the community and will not start prior to 2025

Next steps on the Information Policy

The development of the policy is an **early step in the overall 5-year process.**

Higher
Priority



Lower
Priority

Next steps include:

1. **RFI** on the proposed policy additions and implementation (due March 4)
2. Identify ways to **automate**
3. Provide **further guidance**
4. Provide **training (TOPS)**
5. Provide **support** for adopting open science
6. Support and/or develop technologies

Open-Source Science Initiative @ SMD

Initiates **Transform to OPen Science (TOPS)**, a 5-year program to increase understanding and adoption of open science principles and techniques
Designates **2023 as Year of Open Science**

Continues **investments in open-source science digital infrastructure, cross-divisional AI capabilities and Digital Transformation activities**. (ROSES elements, data catalog, open journal database)

Prototype **common data catalog** by FY22Q4 and expand **Astrophysics Data System**

Initial investments in cross-division **open scientific cloud environments and data analysis platform prototypes**.

\$130M in Divisional investments in Open-Source Science that are **aligned** with this program.

Chief Science Data Officer:
Kevin Murphy

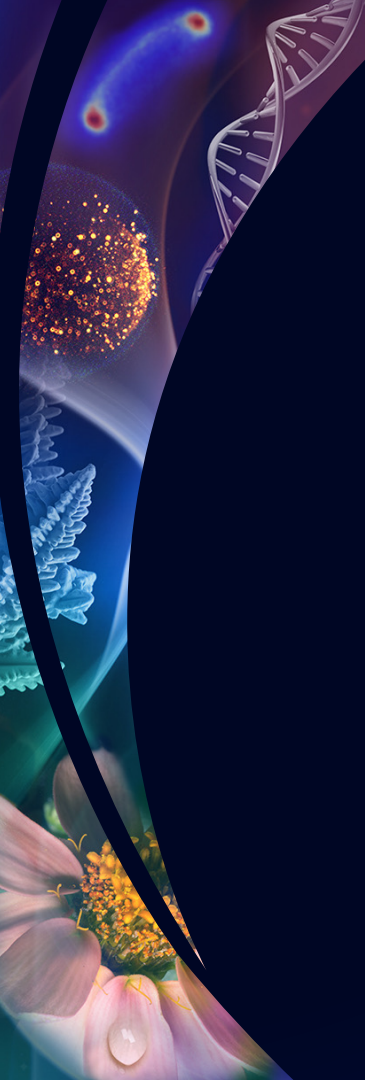
Fiscal year	OSS Total (\$M)
FY21	\$8
FY22	\$21
FY23	\$20
FY24	\$20
FY25	\$20
FY26	\$20
FY27	\$20

Supporting Open Source Software

- ROSES20 E.7 Open Source Tools, Frameworks, and Libraries selected 8 proposals supporting 14 different projects.
- [ROSES21 F.8 Supplemental Open Source Software Awards](#) - Support for existing grant holders to move to open source



Full description of supported projects is available on [NSPIRES](#)



Biological and Physical Sciences Scientific Data Management Policy - *draft*

In guidance with the SMD Policy Document SPD-41, BPS drafted a division specific policy.

- The Biological and Physical Sciences Data Management Policy serves to articulate the governing principles and specify standards of the data and data services.
- These policies are guided by the FAIR principles aimed at maximizing data findability, accessibility, interoperability, and reusability.
- This policy addresses the data lifecycle from measurements to curation, sustainment, and dissemination.
- To reduce the burden on data producers in complying with these policies, the BPS Data Management Environment integrates tools to support the data management processes that create, transform, and transfer to the scientific community at large, various kinds of scientific information.



BPS Data management Policy Draft - Highlights

- **The policies apply to **scientific data** from all BPS-funded activities including:**
 - **Research Data:** Data produced by investigations funded via research awards
 - **Mission Data:** Data produced by NASA programs, projects, or missions that describe conditions and activities relevant to investigations conducted in flight and/or on the ground.
 - **Software:** computer programs in both source and object code that provide users some degree of scientific utility or produce a scientific result or service
- **The policy defines roles and responsibilities for BPS Program Offices, Investigators, Open Science Data Repositories (OSDR), and BPS Projects.**
 - The **OSDR** assure the capability for timely and ongoing delivery of research and mission data by providing guidance and **protected (private/public)** workspace and storage to assure this goal.
 - **Investigators** shall authorize **publication** of research data through the designated OSDR no later than the publication of the peer-reviewed article that describes it
 - **Investigators** shall submit research data throughout the course of the investigation, according to the schedule outlined in the Data Submission Agreement (DSA). These data are to be held by the OSDR and not made public until publication is authorized by the investigator.
 - **Investigators** shall submit all research data mentioned in DSA no later than 1 year following conclusion of grant to the designated OSDR. Independently of peer-reviewed publications status, all data listed in DSA are to be held by the OSDR within this 1-year period, but do not have to be made public for up to 2 years following conclusion of grant, based upon PI's request.
 - **BPS Project** shall be responsible for delivery of mission data to the OSDR no later than 30 days after the completion of data acquisition, calibration, and validation, of agreed upon mission data products with documentation.

Heliophysics Data Policy

- HPD data policy mirrors SMD policy in all respects
- Additional Information and Templates:
 - ROSES Appendix B.1
 - Data Management Plan template found at <https://science.nasa.gov/researchers/templates-heliophysic-division-appendix-b-roses-proposal>
- RFI release on Heliophysics Data Environment: <https://go.nasa.gov/3DxpHPI>



Planetary Science Data Policy

See ROSES-22 Appendix C.1 for updated Data Management Plans & Archiving guidance, which follows SPD-41 policies & definitions

Strongly encourage use of PSD Data Management Plan template

<https://science.nasa.gov/researchers/templates-planetary-science-division-appendix-c-roses-proposals>

Individual PSD program elements may provide instructions that supersede or amplify the requirements described in C.1

Check out the >>new<< Planetary Data website

<https://science.nasa.gov/solar-system/planetary-data-overview>



Progress on Planetary Data Ecosystem (PDE) Independent Review Board (IRB) Recommendations

The final (non-consensus) report delivered in April 2021 and the recordings of the full IRB meetings, are found here: <https://science.nasa.gov/solar-system/pde-irb>

PDE IRB composed of 40 experts from diverse sectors, focus areas, and demographics
Sub-cmtes: Archiving, Searching, Utilization, Mining & Automation, Inter-relational

Report provides 65 prioritized, actionable recommendations for NASA's long-term support of the PDE; many align with NASA's Open Source Science Initiative goals

Updates on PSD's activities to address PDE IRB recommendations:
<https://science.nasa.gov/solar-system/planetary-data-overview>

PDE Progress

Develop the Planetary Data Ecosystem

- PSD launched a [PSD Data Website](#) on science.nasa.gov to provide a centralized location for information on planetary data, the definition of the PDE and identification of its elements, and updates on how PSD is addressing the PDE IRB recommendations.
- The Planetary Data System (PDS) Chief Scientist has been evolved into the [PDE Chief Scientist](#) to increase communication and interoperability of the PDS with the other PDE elements. [Moses Milazzo](#) selected in Dec. 2021 (see [NASA press release](#))
- PSD is planning a virtual [workshop series](#) addressing the PDE IRB recommendations that would be best served by a workshop format, including answering questions such as “What is the PDE?” and “What goes into a Data Management Plan”?

Address Data Preservation & Accessibility Needs

- The Planetary Data System (PDS) is transforming to a more [FAIR](#) archive, e.g., with [cloud migration](#) via PDS Cloud Computing Roadmap and transitioning to a [centralized PDS website](#) ideally with [unified search](#) capabilities across the PDS Nodes.
- PSD is working on [preservation and accessibility](#) of mission-supported laboratory analyses of returned [sample materials](#) for the OSIRIS-Rex mission. AstroMat will be the official laboratory analysis archive for OSIRIS-Rex. PSD is using the OSIRIS-Rex effort to identify and possibly address the broader needs of PSD laboratory sample data curation.

Other

- PSD, as part of SMD AI/ML Initiative, is currently focusing on the development of “benchmark” [training datasets](#) for each of the SMD Divisions and the development of [cross-disciplinary applications](#) of AI/ML models for NASA science.
- Specific [responses](#) from NASA addressing [all](#) of the PDE IRB recommendations (including current status, anticipated timeline to address, and potential future plans) are being provided on the [PSD Data Website](#).

Astrophysics Division Data Policy

“The individual science divisions of SMD may implement additional policies ...”

APD is developing a data policy for their community to provide more guidance on SPD-41.

In developing the policy, APD is has the following credo:

Do no harm

Enable

Incentivize





APD Data Policy

- Have started discussions with stakeholders (internal and external) to clarify requirements of the policy for our Missions and Communities
- Will review RFI responses
- We will keep talking and soliciting feedback on all aspects of data policy: publications, data, and software.
- Early guidance and examples under active discussion
- Terms whose scope is under discussion include:
 - Scientifically or technically relevant data
 - Scientifically or technically relevant software
 - Degree of utility as applied to computer programs
 - Training data used in machine learning



APD Data Policy

The APD is committed to listening to our community:

What would you like to see?

What are you worried about?

What is most important to state?

What should not be stated (to avoid overly constraining implementation and avoid creating non-value-added burdens for compliance)?

Tell us: <https://arc.cnf.io/sessions/r8zx/#!/dashboard> and RFI

Roopesh.Ojha@nasa.gov

Open-Source Science Policy for Earth System Observatory

- A. All mission data, metadata, software, databases, publications, and documentation shall be available on a full, free, open, and unrestricted basis starting in Phase B with no period of exclusive access.
- B. Science workshops and meetings shall be open to broad participation and documented in public repositories.

1 Software shall be developed openly in a publicly accessible, version-controlled platform using a permissive software license allowing for community use and contributions.

2 Manuscripts shall be published with open access licenses; versions of as-accepted manuscripts shall be made available as open preprints and deposited in a NASA or [Partner] repository upon publication.

3 All mission data, calibration information, and simulated products supporting development and validation of algorithms shall be made available without any conditions to use.

4 Scientific data, metadata, software, publications and documentation shall be archived and made available by NASA and/or [Partner] starting in Phase B.

5 NASA and [Partner] software, documentation and data shall be properly marked, cited, and/or attributed. Metrics to measure and acknowledge open-source science contributions will be developed.

6 NASA and [Partner] will mutually develop an Open-Source Science Plan that specifies details of collaboration.

Envisioning DAACs as Science Enabling Centers

1

Data and metadata stewardship

Ensuring **quality and fitness for purpose** of the organization's data and metadata assets

2

Information management

Capturing and cataloging scientific information from publications into **searchable databases linked to data resources**

Methods: semantic search, natural language processing, machine learning techniques

3

Open-source software support

Managing and supporting open-source software development projects: cataloging, documentation, review for long-term sustainability, and user support

4

Cross-mission science and modeling

Addressing cross-mission and division science as well as dedicated support for modeling communities to better integrate and fuse observational data into models

5

User support

First and second tier of **support for user community (scientists and nonscientists)**

Capable of answering scientific questions about data and information, evaluating and merging community contributions to open-source software, addressing technical questions, and developing unique software to serve their science communities.



Open-Source Science for Earth System Observatory Mission

- **Goal:** Identify and assess potential architectures that can meet the ESO mission science processing objectives, enable data system efficiencies, promote open science principles, and seek opportunities that support Earth system science.
- **ESO Processing Workshop #1 (Oct 19-20, 2021)** focused on gathering needs for evaluating open science data system architectures to support Earth system science and mission data system efficiencies

YOU'RE INVITED!

ESO Processing Workshop #2: March 1-2 2022

Open invitation for agencies to share insights into their data processing approaches and share state of the art practices in big data processing systems.

Transform to Open Science

Accelerating Scientific Discovery

These activities are designed to **support and strengthen** other NASA SMD initiatives on Inclusion, Diversity, Equity, and Accessibility (IDEA) and work for environmental justice.

> PROTECTING & IMPROVING LIFE ON EARTH
> LIFE ON OTHER PLANETS
> MYSTERIES OF THE UNIVERSE

Overview

- TOPS 5-year initiative will act as a catalyst to **jump-start** a suite of coordinated activities designed to rapidly transform science.
- Designate **2023 as the Year of Open Science (YOOS)**.
- Our focus for TOPS will be on early career scientist in NASA SMD communities, and welcome participation and coordination with other groups.

In 5 years, TOPS will:

1. **Increase understanding and adoption of open science principles and techniques in our Mission and Research Communities:** 20K scientists earn open science certifications, achievements, and badges at summer schools, society meetings, & other events
2. **Accelerate major scientific discoveries through supporting the adoption of open science:** ROSES solicitation to support major scientific discoveries using open science methods in each division: 5 major results within 5 years
3. **Broaden participation by historically excluded communities:** Double participation by historically excluded communities in submitted proposals, applications from students, and participation in mission teams.



Public Comment on the Policy

RFI due March 4, 2022. Information requested on:

- How will the proposed changes to the existing information policy impact the research activities of your communities?
- What support, services, training, funding, or further guidance is needed to support the successful implementation of the existing or proposed information policy?

Questions can also be sent to HQ-SMD-SPD41@mail.nasa.gov

The background of the slide is a cosmic scene. The top half features a dark blue and black space filled with numerous small stars and a prominent, bright blue nebula on the right side. The bottom half transitions into a warmer color palette, with a golden-yellow and orange glow on the left, transitioning into a green and blue glow on the right, also filled with stars and nebulae. A light blue horizontal band is centered across the image, containing the text.

Back Up slides

OPEN SCIENCE

Open science is a collaborative culture enabled by technology that empowers the **open sharing of data, information, and knowledge** within the scientific community and the wider public to accelerate scientific research and understanding.



OPEN-SOURCE SCIENCE

Build on concepts from Open-Source Software, expanding participation in developing code, applying to the scientific process to accelerate discovery by openly conducting science from project initiation through implementation.

WHAT IS “OPEN” ABOUT OPEN-SOURCE SCIENCE?

The Four Meanings of “Open” in Open-source Science



OPEN (**TRANSPARENT**) SCIENCE
scientific process and results
should be visible, accessible, and
understandable

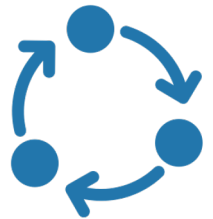


OPEN (**INCLUSIVE**) SCIENCE
process and participants should
welcome participation by
and collaboration with diverse
people and organizations

OPEN (**ACCESSIBLE**) SCIENCE
data, tools, software,
documentation, and
publications should be
accessible to all (FAIR)



OPEN (**REPRODUCIBLE**) SCIENCE
scientific process and results
should be open such that they
are reproducible by members of
the community



Expanding access to publications

- Encourage and support publishing as full open access and the use of preprint services.
- Expanding the Astrophysics Data Service (ADS) to include other SMD divisions to enable search and connections between publications, data, and software. This includes improving coverage of NTRS.
- SMD awarded a 3 year grant to arXiv.
- Supporting the Space Act Agreement with CHORUS by STI for improved tracking of publications and automated access to publications.
- Providing improved guidance to the community about how to make their publications accessible.

Variations

SPD-41a also includes a process for variations.

Variations:

- **are meant to be rare**
- **can be requested, for example where the policy may cause an undue burden.**

As part of the process, there can be a conversation about what steps might help mitigate the process or to help make the information more accessible or what parts of the policy the variance is applicable to.

Questions about variations can be sent to HQ-SMD-SPD41@mail.nasa.gov, and requests for variations can be made through your program officer.