National Aeronautics and
Space Administration

EXPLORE SOLAR SYSTEM&BEYOND

Update on the Planetary Data Ecosystem (PDE)

Becky McCauley Rench

PAC Meeting

June 23, 2022

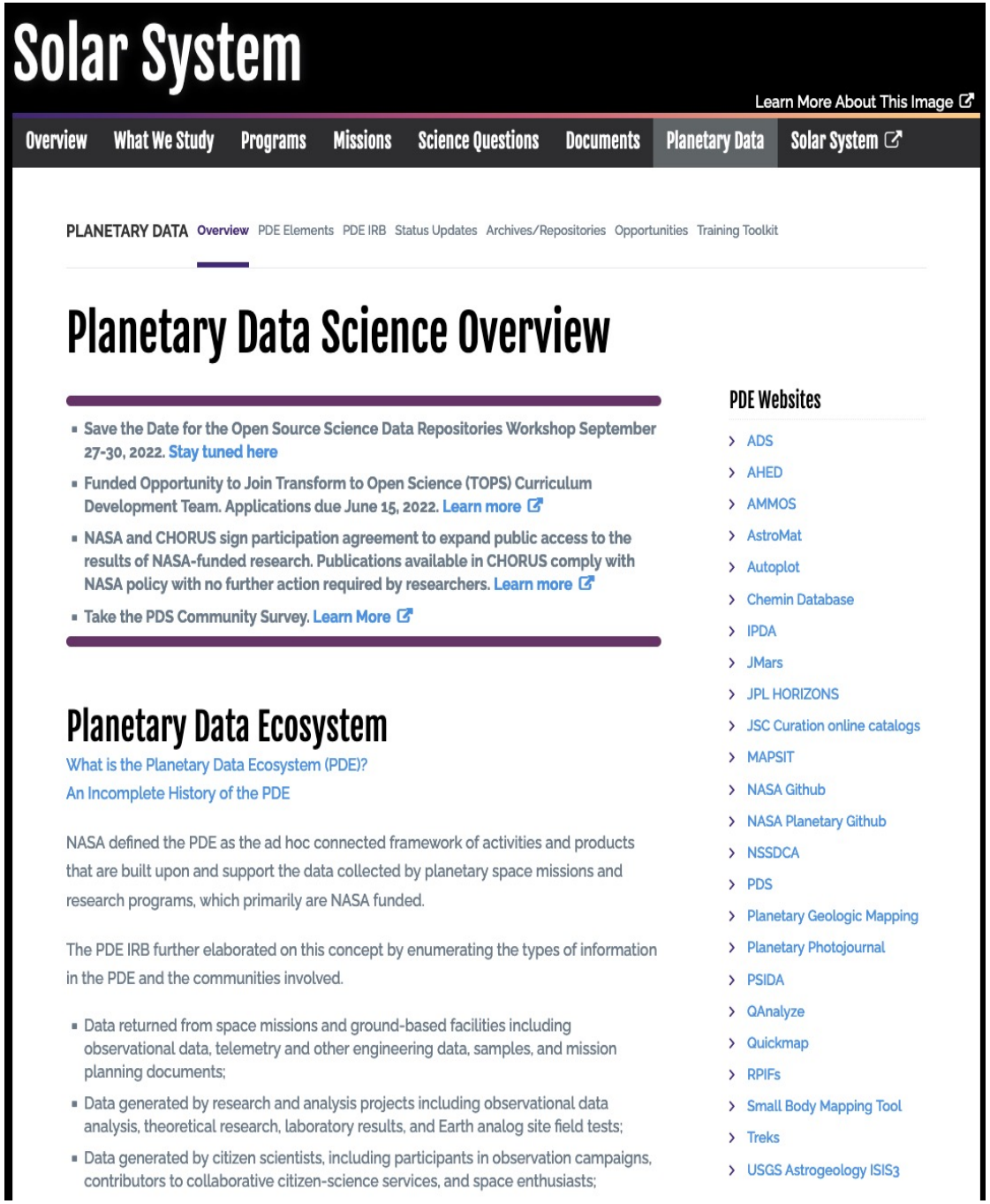
Timeline & Schedule

- April 22, 2021 – NASA received the PDE IRB report
- July 2021 – PDS Discipline Node Programmatic Review, which was structured to address some PDE IRB recommendations
- December 2021 – PSD Data webpage on science.nasa.gov launched; PDE Chief Scientist announced
- April 2022 – PDS Support Node Programmatic Review, which was also structured to address some PDE IRB recommendations
- June 2022 – Selection of TWSC proposal to support Planetary Data Training (PI: David Williams, ASU)
- **Future Plans:**
  - Summer/Fall 2022 – Specific response from NASA addressing ALL recommendations to include current status, anticipated timeline to address (if applicable), and potential future plans on PDE webpage (responses to be prioritized at end of this deck and to be posted on PDE webpage)
  - Late 2022/Early 2023 – PDE Workshop Series, Identification of internal Planetary Data Officer, Release of Planetary Information Policy (will build on SPD-41)

PDE Webpage Launched

- Includes announcements relevant to the planetary community
- Point of access for various elements of the PDE
- Status updates on PSD progress in responding to PDE IRB recommendations
- Includes list of archives & repositories that are long-term and meet FAIR guidelines
- List of community identified data needs and available trainings

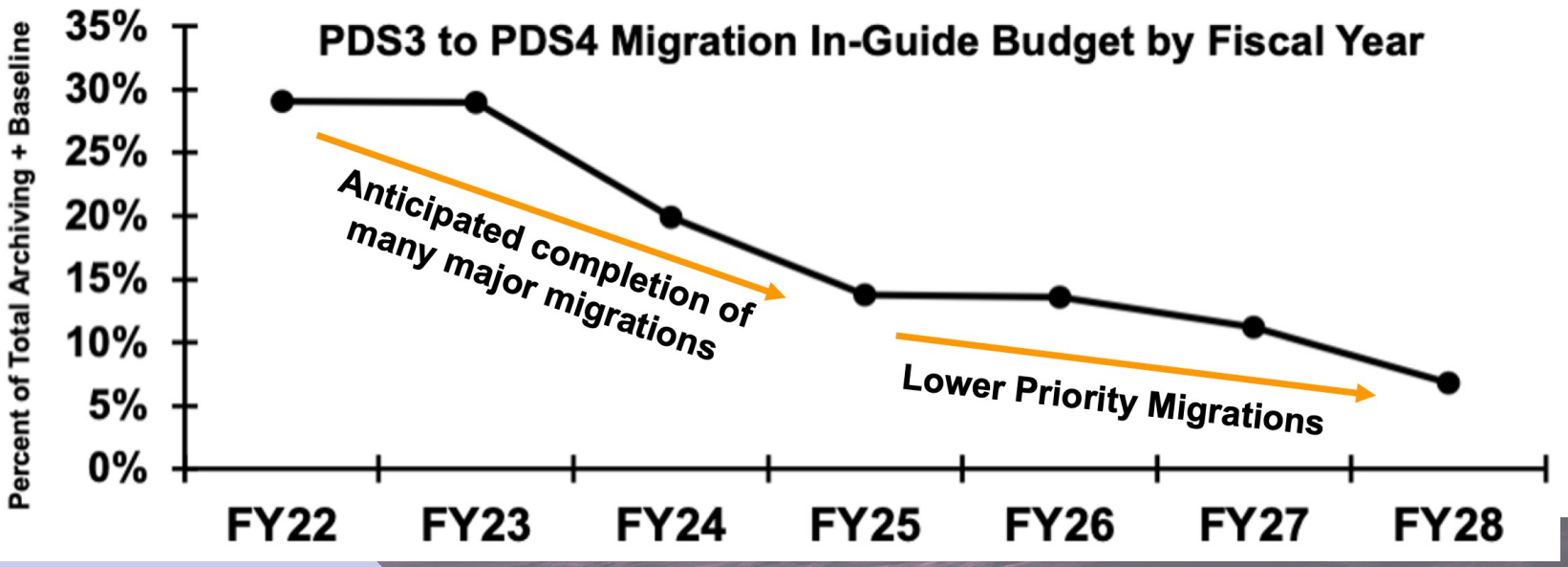
See something that should be added?
Email hq-pde@mail.nasa.gov



Improving the PDS – PMSR22 & PDS4 Migration

Between additional funding to PDS and moving active missions toward PDS4 deliveries*, PDS is on track to complete major migrations to PDS4 by FY25 and completion of PDS4 migration by FY28.

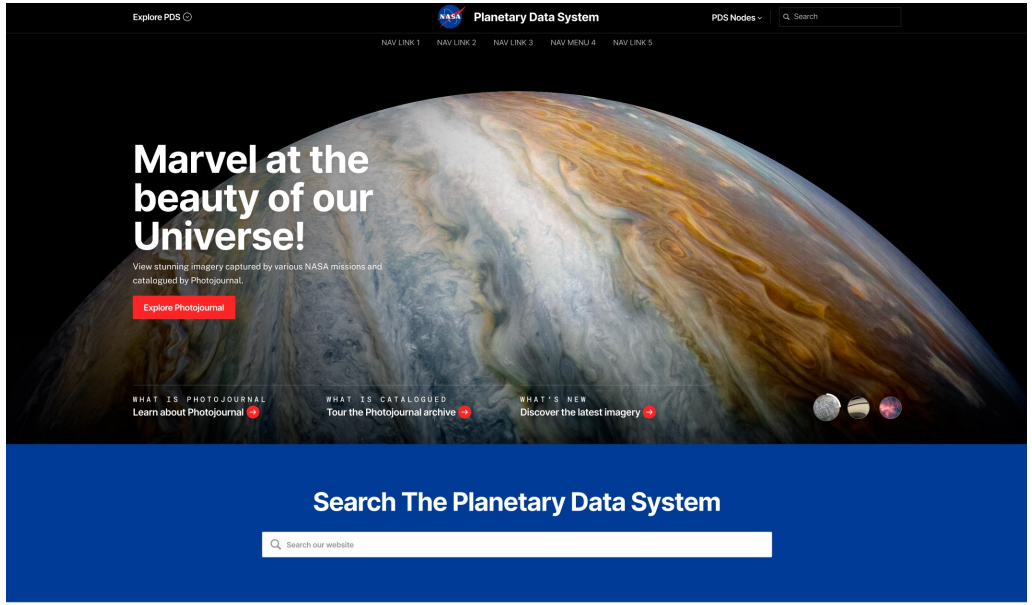
*Recent PMSR required extended missions to start delivery in PDS4 for the extended mission phase. Back conversion of PDS3 deliveries to PDS4 were proposed for some missions as overguides. These two activities are synergistic and overguides for back-conversion of PDS3 to PDS4 were approved for MSL and MRO.

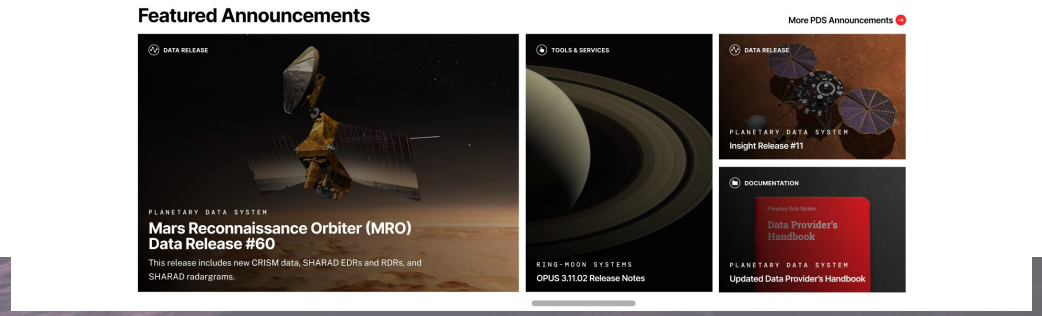


Improving the PDS – Web Modernization

To provide the current and future planetary science community with an **approachable** and **unified PDS web presence**, **user experience**, and **portal to the data**, **documentation**, and **tools** necessary for **<u>enabling scientific discovery.</u>**

- Providing a **consistent, unified, modernized** PDS website
- Improving **discoverability** of data, documentation, tools and information across all internationally-sponsored planetary science archives.
- Improve **user experience** and **user satisfaction**
- **Integrate** with other NASA and international web sites





Planetary Data Training Workshops!

- In June 2022, PSD selected a proposal (PI: David Williams, ASU) to offer a series of Planetary Data Training Workshops starting in late 2022/early 2023.
- Workshops will focus on planetary data management, planetary Geographic Information Systems (GIS) training (ArcGIS, open source GIS, and JMARS), ISIS3 for image processing, and SOCET SET-Ames Stereo Pipeline for digital elevation model (DEM) production.
- Workshops will be open to all members of the U.S. planetary science community, from undergraduate and graduate students to postdocs and early career scientists to established professional scientists.
- In the interests of supporting diversity, equity, and inclusion, two travel grants FOR EACH EVENT would be funded to support participation by planetary scientists from underrepresented communities.

Advancing Data Ecosystems, Together

PSD is leveraging SMD initiatives to advance our PDE goals

- AI/ML Training Dataset Pilot: PDS RMS Node will leverage funding from the SMD AI/ML Initiative to develop a benchmark training dataset for planetary science as well as develop a pilot ML-driven classification and search capability
- Cloud-Ready Data Pilot: USGS will leverage SMD cloud computing resources and funds to convert HiRISE RDRs from PDS3/Jpeg2000 to Cloud Optimized GeoTIFFs (COGs)
- PDS moving towards a cloud environment facilitated by SMD core data services

Progress in Responding to PDE IRB Recommendations

- Develop the Planetary Data Ecosystem
  - PROGRESS & CURRENT ACTION SUMMARY: PSD selected a PDE Chief Scientist, launched a PDE webpage, and completed the programmatic reviews of the PDS Nodes. PSD supported the development of SPD-41, with a PSD Information Policy to be available by ROSES-2023. Additionally, PSD is planning to recruit a Planetary Data Officer, and initiate a PDE workshop series. More details in backup slides.
- Address Data Preservation Needs
  - PROGRESS & CURRENT ACTION SUMMARY: PSD is working with the SMD Open Source Science Initiative to advance a variety of data needs of the community, both as identified in the PDE IRB report and as part of SPD-41 implementation. Additionally, PSD currently supports radar data analysis, publication, and archiving of Arecibo data at the PDS SBN and has added a PDS Radio Science sub-Node that includes a Planetary Radar Advisory Role. Lastly, the PDE webpage includes information about Community-Identified Data Needs, which can be delivered to NASA via the AGs. See next slide on Astromat and more details in backup slides.
- Address Barriers to Data Use and Development
  - PROGRESS & CURRENT ACTION: PSD is supporting Planetary Data Trainings, supporting investigation of a PDS Engineering Data Node at JPL, working with the SMD Open Source Science Initiative to identify core services and leverage assets across SMD, piloting examples of cloud-ready and analysis-ready data sets, and investigating appropriate review cycles and criteria for PDE elements.
- Other Recommendations
  - PROGRESS & CURRENT ACTION: Specific response from NASA addressing ALL recommendations to include current status, anticipated timeline to address (if applicable), and potential future plans will be finished in Summer/Fall 2022 and posted on the PDE webpage.

Address Data Preservation Needs - Astromat

PROGRESS & CURRENT ACTION: PSD is actively working preservation of mission-supported laboratory analyses of returned sample material for the OSIRIS-REx mission. PSD has been working with the Astromaterials Data System (Astromat) via the Johnson Space Center Astromaterials Acquisition & Curation Office to address this need. Astromat is currently completing a special study to lay out the implementation plan for appropriate archiving of OSIRIS-REx, and other returned sample missions, laboratory sample data curation. Astromat is actively working with the PDS to determine interoperability. Two fundamental principles are that archived mission-supported laboratory analyses of returned samples must be archived at a high quality (e.g. PDS4 compliant) and that users should be able to access all data archived for a particular sample in one place (e.g. interoperability).

In response to PDE IRB Recommendation:

R33: NASA should establish a requirement for the preservation of mission-supported laboratory analyses of returned sample material that makes the information accessible to the planetary science community. Time is of the essence to establish these requirements, as NASA will receive the largest sample return since Apollo in approximately two years.

R34: NASA should require data preservation with appropriate metadata in an approved archive or repository for data produced by laboratory analysis of returned samples supported by ROSES Data Analysis Programs (DAP).

ADDITIONAL SLIDES IF TIME ALLOWS

PDE Workshop Series: In Progress

- PSD plans to address some of the PDE IRB recommendations using virtual workshops to start the conversation and further define planetary science community needs.
- PSD has identified 4 tentative workshop topics, with more to be added:
  1. What is the PDE?
  2. What goes into a Data Management Plan?
  3. PDE User Community Needs/Browsing through the PDE?
  4. What are analysis-ready data sets?

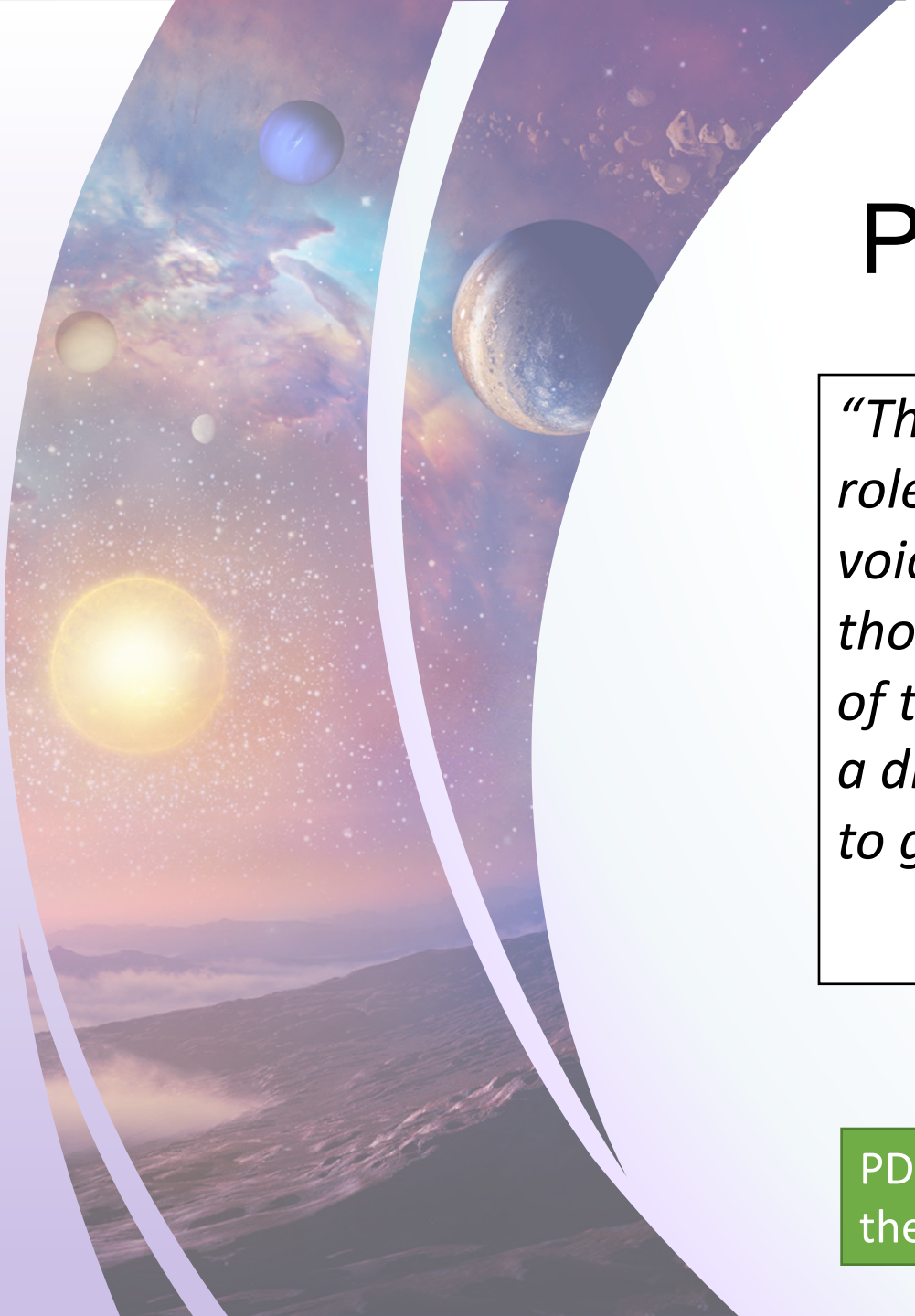
Other Updates Related to PDE

PDAR(T)

- Revamped PDAR(T) still in formulation and to be released in ROSES-2023
- The hiatus on tool development PDAR(T) proposals is not slated to last longer than one year and should be significantly improved

Mission Software Post-Mission

- PSD is strategizing on best practices for capturing software developed for missions, post-mission.
  - Example: JMars currently supported by Odyssey
  - Example: Quickmap currently supported by LRO



PDE Chief Scientist

"The establishment of this independent chief scientist role provides the planetary data community with a voice directly representing it to NASA. This ensures that those who use and rely on the PDS and other elements of the ecosystem have a direct link to NASA, so we hear a diversity of input and experience as the PDE continues to grow and evolve."

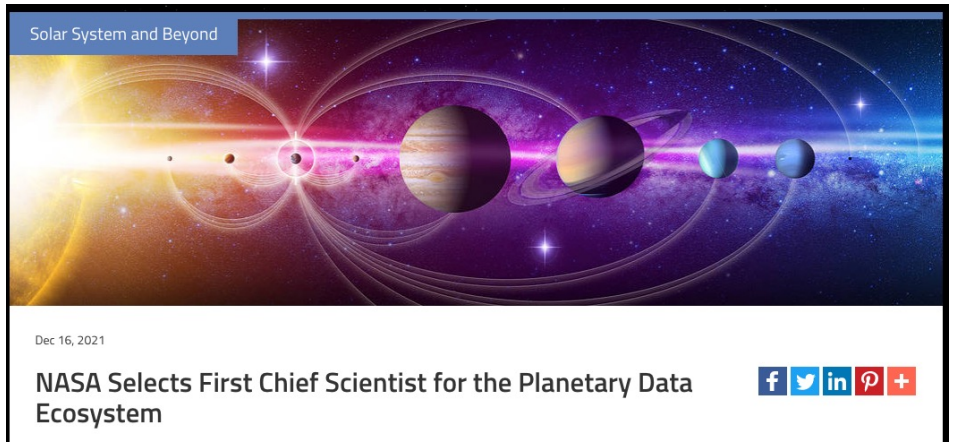
– Lori Glaze

PDE Chief Scientist Moses Milazzo will be presenting next on the Agenda

BACKUP: MORE DETAIL ON PDE IRB RECOMMENDATION RESPONSES

Develop the Planetary Data Ecosystem

PROGRESS & CURRENT ACTION: NASA PSD selected a PDE Chief Scientist, Moses Milazzo, in December 2021. The PDE Chief Scientist provides an independent link between the larger PDE community, the Planetary Data System (PDS) and NASA Headquarters, and also refines and represents the PDE to NASA. PSD is also working to develop a PDE workshop series. Once the PDE workshop series takes off, we will use that series as an opportunity to investigate a larger community-led group. The final workshop in the PDE workshop series could be utilized to create such a community group. The workshop topics are being drafted by the PDE working group. Once ready, the info will be presented to LPI to determine whether and how LPI could host the workshop series.

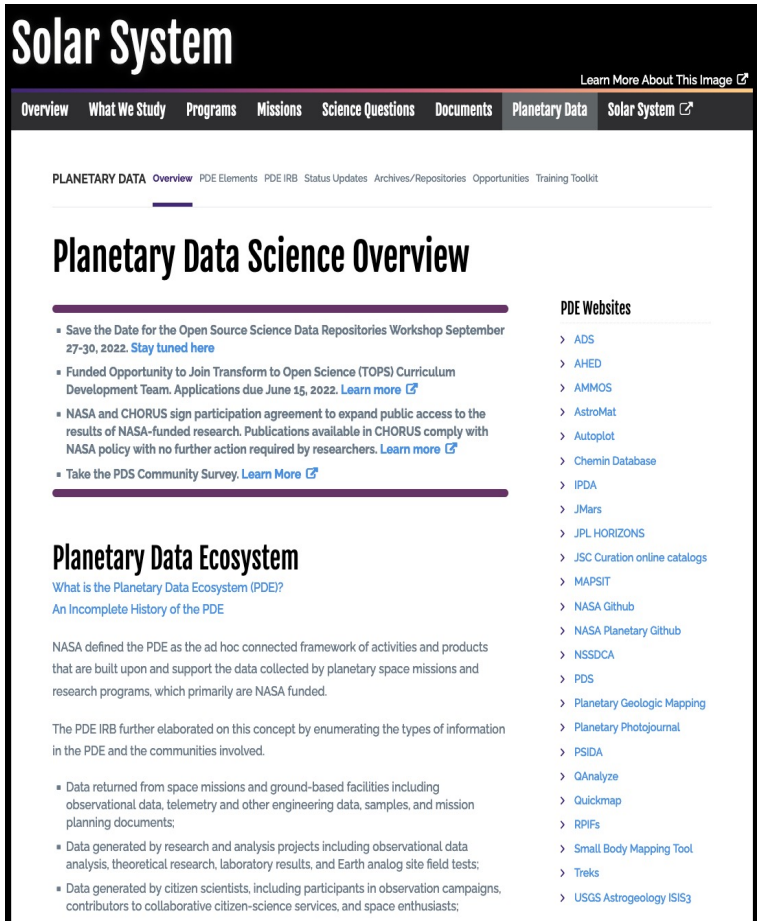


In response to PDE IRB Recommendation:

R4: NASA should ensure that a sustained, community-led coordinating organization for the PDE exists that mirrors the other Planetary Assessment or Analysis Groups (AGs), reports to the Planetary Science Advisory Committee, and meets regularly. (Non-consensus)

Develop the Planetary Data Ecosystem (cont.)

PROGRESS & CURRENT ACTION: The Planetary Data webpage was launched on science.nasa.gov in December 2021. The content for this webpage was seeded by the PDE IRB report and has been updated based on input from planetary community members. The webpage can serve as a portal to all the identified elements of the Ecosystem. As part of NASA's web modernization efforts, the future location of the content on this webpage may migrate and be improved upon as a more user-friendly portal to entering the Ecosystem. PSD is also working to develop a PDE workshop series. The PDE workshop series is slated to have topics in this area – “What is the PDE?” and “PDE User Community Needs/Browsing through the PDE?”. The workshop topics are being drafted by the PDE working group. Once ready, the info will be presented to LPI to determine whether and how LPI could host the workshop series. Once established, it is expected a PDE community-led group will continue to provide information on the state of the PDE.



In response to PDE IRB Recommendations: R01 and R02 (see next slide for details)

Develop the Planetary Data Ecosystem (cont.)

Previous slides Progress & Current Actions are in response to PDE IRB Recommendation:

R1: NASA should proceed with developing the concept of the Planetary Data Ecosystem so that the usability and archival needs of the entire planetary sciences community—all people, professional or amateur, who produce, provide, and/or use data—are better met.

R2: NASA should lead work to refine the full scope of the Planetary Data Ecosystem and build community consensus around the Ecosystem. NASA should continue to refine the short definition as well as the detailed list that answers the question: "What is the PDE?" that clearly differentiates it from the PDS.

Develop the Planetary Data Ecosystem (cont.)

PROGRESS & CURRENT ACTION: PSD is working to group elements of the PDE within a managed portfolio. This will assist in completing regular reviews that tackle concepts of appropriate governance and service levels. PSD is going to recruit a Planetary Data Officer that will assist in gathering information on Ecosystem elements and conducting regular reviews of the PDE elements and how they fit together. The Planetary Data webpage and PDE workshop series will also provide information and context for this work. Once established, it is expected a PDE community-led group will continue to provide information on the state of the PDE.

In response to PDE IRB Recommendation:

R3: NASA should ensure that the responsibilities, accountabilities, governance, and service levels for those elements of the Ecosystem that are funded by the NASA Planetary Science Division are clearly defined.

Develop the Planetary Data Ecosystem (cont.)

PROGRESS & CURRENT ACTION: The first priority of the PDS is to be the long-term archive of digital data products returned from NASA's planetary missions, and from other kinds of flight and ground-based data acquisitions, including laboratory experiments. Building on that, the second priority of the PDS is to ensure the usefulness and usability of that data for the planetary science community. The PDS Discipline and Support Nodes completed their Programmatic Reviews in 2021 and 2022, respectively. The outcome of each of the reviews has been to lean forward and the PDS Nodes have been funded at their requested levels. Included in the description of Node responsibilities is the data scope of each Node, which is consistent with the description of each Node on the PDS webpage. The centralized web presence of the PDS, slated to take place by the end of 2024, will improve the discoverability and understanding of its role in the Ecosystem. Additionally, the new PDE webpage assists in making the other components of the Ecosystem more widely known, thereby helping to make clear the specific role of the PDS within the PDE. Lastly, the new SPD-41 (SMD Information Policy) articulates what NASA's expectation is for data preservation and distribution. PSD's Information Policy will build upon SPD-41 and is in work and will be made public with ROSES-2023.

In response to PDE IRB Recommendation:

R14: Consideration should be given to how to make clear the differing responsibilities and expectations of the data preservation mission from the distribution of usable data. Consistent with Recommendation 2 for the broader Ecosystem, the prioritize goals and scope of PDS need to be carefully and explicitly defined by NASA, with input from the Ecosystem and broader community, and clearly articulated to all members of the community. Mandates above and beyond the agreed-upon scope must be negotiated and accompanied by commensurate funding. NASA should fund PDS nodes at levels appropriate to the full scope of work defined by the selected proposals as well as any accumulated duties.

Address Data Preservation Needs

PROGRESS & CURRENT ACTION: PSD currently supports radar data analysis, publication, and archiving of Arecibo data at the PDS SBN. Meetings between the PDS SBN and the Arecibo, JPL, and Goldstone radar groups to coordinate formats and processes among their substantial radar data archiving efforts are ongoing and continue to be of priority. SBN is actively working with these radar observers to prepare and submit their data to the PDS and expand SBN holding of ground-based radar data observations. Additionally, PSD has supported the creation of a Radio Science sub-Node of the PDS established in collaboration with the Planetary Radar and Radio Sciences Group (PRRSG) at JPL and which provides a Planetary Radar Advisory Role to the PDS. No further actions are planned to be taken in response to this recommendation, as the remaining component of archival of software is part of a larger Ecosystem discussion.

In response to PDE IRB Recommendation:

R31: NASA should establish an archive for planetary radar data either within the PDS Small Bodies Node or separately. This archive should facilitate preservation and usability of data at all processing levels by preservation of data processing procedures (or software). Because of the unique situation of Arecibo Observatory, time is of the essence to preserve the data and prevent irretrievable loss.

Address Data Preservation Needs - Astromat

PROGRESS & CURRENT ACTION: PSD is actively working preservation of mission-supported laboratory analyses of returned sample material for the OSIRIS-REx mission. PSD has been working with the Astromaterials Data System (Astromat) via the Johnson Space Center Astromaterials Acquisition & Curation Office to address this need. Astromat is currently completing a special study to lay out the implementation plan for appropriate archiving of OSIRIS-REx, and other returned sample mission. Astromat is actively working with the PDS to determine interoperability. Two fundamental principles are that archived mission-supported laboratory analyses of returned samples must be archived at a high quality (e.g. PDS4 compliant) and that users should be able to access all data archived for a particular sample in one place (e.g. interoperability).

In response to PDE IRB Recommendation:

R33: NASA should establish a requirement for the preservation of mission-supported laboratory analyses of returned sample material that makes the information accessible to the planetary science community. Time is of the essence to establish these requirements, as NASA will receive the largest sample return since Apollo in approximately two years.

R34: NASA should require data preservation with appropriate metadata in an approved archive or repository for data produced by laboratory analysis of returned samples supported by ROSES Data Analysis Programs (DAP).

Address Data Preservation Needs (cont.)

PROGRESS & CURRENT ACTION: The PDE IRB report identified an initial list of data preservation needs. One response to those identified needs has been to support the development of a modeling annex at the PDS Atmospheres Node. Additionally, SMD released an RFI to comment on SPD-41. RFI responses could include information “about what support, services, training, funding, or further guidance is needed to support the successful implementation of the existing or proposed information policy.” This has been one recent avenue for community needs to be identified. Lastly, the PDE webpage includes a section for Community-Identified Data Needs and states that AGs can provide this information to NASA HQ. The AGs were specifically asked for this information in Fall 2021. The PDE working group will continue to craft strategies to identify and prioritize the needs of the planetary science community, which may include connections to the revamped PDAR(T) in ROSES-2023.

In response to PDE IRB Recommendation:

R28: NASA should establish a carefully crafted strategy to identify and prioritize the data preservation needs of the planetary science community that are not currently being addressed.

Address Data Preservation Needs (cont.)

PROGRESS & CURRENT ACTION: The PSD PDE working group completed an internal, informal survey of non-PDS archives supported by PSD and found most to be non-FAIR compliant. This informal survey also identified that archives to support model output and other large/complex derived data products were rare. Many PIs are currently using supplemental material in publications, and institutional archives to host data and codes. The core of this recommendation spans beyond PSD and into SMD. As such, PSD will continue to work within the SMD Open Source Science Initiative to progress in this area. Addressing the requirements of the new SMD Scientific Information Policy (SPD-41) will support FAIR principles for PSD data. Currently, non-PDS archives supported by the PSD are included on the PDE webpage. The PDE working group will continue to gather information on what archives and repositories are currently used by the planetary community and whether they meet the standards outlined in SPD-41 and the forthcoming Planetary Information Policy.

In response to PDE IRB Recommendation:

R29: NASA should consider ways of archiving outside of the PDS that are amenable to creating FAIR and standards-based archives of these growing data sets.

Address Barriers to Data Use and Development

PROGRESS & CURRENT ACTION: While the existing process for mission data archival meets NASA's requirements, PSD is keenly aware that regular internal assessment of a mission's Data Management Plan is necessary as our understanding of complex instruments and mission goals advance. The PDE working group will continue to consider how to best address this recommendation.

In response to PDE IRB Recommendation:

R21: NASA should treat mission data archival as a systems engineering concern by including early funding for mission data acquisition, processing, and archiving of data and foundational data products (including cartographic products, data acquisition contextual information, coordinate system standards, etc.) so that they are planned well in advance of data acquisition.

Address Barriers to Data Use and Development (cont.)

PROGRESS & CURRENT ACTION: In June 2022, PSD selected a proposal (PI: David Williams, ASU) to offer a series of Planetary Data Training Workshops starting in late 2022/early 2023. Workshops will focus on planetary data management, planetary Geographic Information Systems (GIS) training (ArcGIS, open source GIS, and JMARS), ISIS3 for image processing, and SOCET SET-Ames Stereo Pipeline for digital elevation model (DEM) production. Additionally, we have created space on the PDE webpage to highlight training opportunities. PSD expects to continue supporting training opportunities.

In response to PDE IRB Recommendation:

R23: NASA should provide regular, accessible, and effective training programs for researchers, data producers, mission specialists, and others who need to archive with the PDS. This should not just be provided by the PDS: entities with experience delivering to the PDS should also be involved. There should also be training for peer-review of data archives. We also recommend that this training and documentation address data preparation from the perspective of reusability and interoperability, such as the Earth Science Data Systems Working Group (ESDSWG) Data Product Development Guide (DPDG) for Data Producers.

R64: NASA should seek to expand opportunities for intermediate to advanced technical training in topics related to accessing, using, and processing planetary data.

Address Barriers to Data Use and Development (cont.)

PROGRESS & CURRENT ACTION: The PSD PDE working group is identifying possible actions to be taken at NASA HQ in this area. A few relevant actions include developing a PDE workshop series (R50), supporting investigation of a PDS Engineering Data Node at JPL (R52), working with the SMD Open Source Science Initiative to identify core services and leverage assets across SMD (R05), piloting examples of cloud-ready and analysis-ready data sets (R52), and investigating appropriate review cycles and criteria for PDE elements (R11).

In response to PDE IRB Recommendation:

R50: NASA should develop outreach to user communities within the Planetary Data Ecosystem, assess user needs, and develop focused educational and documentation materials that meet highest-priority needs.

R52: Relevant elements of the Ecosystem should support the delivery of higher-level and analysis-ready data products in well-documented and broadly used protocols and formats, even where those formats might not be appropriate for primary data. This should include broadening support across the Ecosystem for a wider variety of data and information formats, such as engineering data; data models; sound and imaging data; and physical collections attached to planetary missions.

R05: NASA should expand intra- and inter-agency efforts to ensure that best practices, lessons learned, and appropriate technologies are shared and implemented across Planetary Data Ecosystem elements.

R11: The Planetary Data Ecosystem should regularly (on a one- to two-year time scale) assess the Findability, Accessibility, Interoperability, and Reusability (FAIR) of data across each PDE element for machine-actionable access to data. This assessment should be used to establish the priorities for Ecosystem management and advisory groups.

PLANETARY DATA ECOSYSTEM CHIEF SCIENTIST REPORT

Planetary Data Ecosystem Report

Presented by
Moses Milazzo (he/him)
PDE Chief Scientist

Planetary Data Ecosystem Independent Review Board

PLANETARY DATA ECOSYSTEM CHIEF SCIENTIST REPORT

Presented by
Moses Milazzo (he/him)
PDE Chief Scientist

The Planetary Data Ecosystem (PDE) Independent Review Board (IRB) Reported in April 2021:

- Recommendation 4: “NASA should hould ensure that a sustained, community-led coordinating organization for the PDE exists that mirrors the other Planetary Assessment or Analysis Groups (AGs), reports to the Planetary Science Advisory Committee, and meets regularly.
- NASA Announced selection of PDE Chief Scientist (16 Dec 2021).
- CS work began late winter / early spring.
- CS is an Independent Contractor, not a NASA employee

The final report is available at:
https://science.nasa.gov/researchers/science-data
(scroll to the bottom on the page)

PLANETARY DATA ECOSYSTEM CHIEF SCIENTIST REPORT

(Incomplete) History of the Planetary Data Ecosystem

1963 – USGS Astrogeology Science Center established

1966 – NASA Space Science Data Coordinated Archive (NSSDCA) established

1977 – NASA Regional Planetary Image Facilities (RPIFs) established (sunset in 2020)

1982 – National Academy of Sciences Committee on Data Management and Computation (CODMAC) chartered

1985 – NASA's Advanced Multi-Mission Operations System (AMMOS) initially developed

1989 – NASA Planetary Data System (PDS) established

1998 – NASA Center for Near Earth Object Studies (CNEOS)* established

2005 - NASA Data Analysis Program (e.g. DDAP, MDAP) established

2014 – NASA Planetary Data Archiving, Restoration, and Tools (PDART) program created

2014 – Mapping and Planetary Spatial Infrastructure Team (MAPSIT) established

2016 – Idea of Planetary Data Environment/Ecosystem starts to take hold at NASA HQ

2021 – Planetary Data Ecosystem (PDE) Independent Review Board (IRB) report

The Planetary Data Ecosystem today

PLANETARY DATA ECOSYSTEM CHIEF SCIENTIST REPORT

Physical Facilities	Archives	Online Repositories, Registries, and Portals	Public Communications
• NASA Center Archives • Spectroscopy Labs • Astromaterials Acquisition and Curation	• PDS • CNEOS • HITRAN • IRSA • MAST • NSSDCA • STI • SPDF • ESA's PSA	• AstroMat • ADS • Institutional Repositories • AHED • NASA Exoplanet Archive • NASA Open Data and Software Portal • Scientific Journals	• Active Mission Websites and Communications • Raw Images from Active NASA Missions • Instrument-specific Websites • NASA Photo Galleries • Science Nuggets • Treks • News Media Information
Sources of New Data	**Data Standards**	**Software**	**K-99+ Education Portals**
• Missions • DAPs • PDAR(T) • R&A • Scientists • NASA's • Domestic and International Partners	• PDS • FGDC • IAU • ISO • FAIR (access) • Open (access) • …	• AMMOS • Autopilot • ISIS3 + 4 • JMARS • SBMT • ASP • …	• Challenger Centers • NASA Kids' Club • NASA Science Space Place • NASA STEM Engagement • Universities • Junior and Comm Colleges

Core Values

The IRB developed a set of core values

- **First, do no harm:** Avoid the law of unintended consequences.
- **FAIR:** Facilitate participation in the PDE by adhering to FAIR data principles of Findability, Accessibility, Interoperability, and Reusability.
- **Open:** Advocate open science practices, including open access, open data, open code, open software/tools, and others.
- **Collaborative:** Encourage international collaboration. Welcome new participants from both inside and outside the professional space exploration community.
- **Effective:** Provide timely, useful support to user communities, especially data producers.
- **Practical:** Pursuit of ideal solutions may sometimes leave the Ecosystem with no solution at all rather than a solution that is sufficient.

Planetary Data Ecosystem Chief Scientist Report

Presented by Moses Milazzo
PDE Chief Scientist

The PDE IRB report is available at:
https://science.nasa.gov/researchers/science-data
(scroll to the bottom on the page)

67 findings and 65 recommendations were organized into 5 themes:

- The Planetary Data Ecosystem Concept
- Planetary Data Stewardship
- Systemic Barriers to Data Preservation
- Barriers to Access and Usability
- Barriers to Development

+ a concluding “Pathway Toward an Ideal State” section

Planetary Data Ecosystem Independent Review

The highest priority recommendations fell into three groups:

- Develop the Ecosystem
- Address Data Preservation Needs
- Address Barriers to Use and Development

Develop the Ecosystem

Develop the Ecosystem

Plain language findings and recommendations summary

- PDE is a good idea and should be formalized
- PDE ≠ PDS; it is much broader
- Lack of communication among PDE elements causes inefficiencies and data losses
- NASA is an important leader, but it could lead more effectively by participating more in established communities
- NASA needs to learn from non-planetary communities to increase accessibility and use of planetary data

Develop the Ecosystem High Priority Recommendations

Group 1: Develop the Ecosystem

- ✓ Establish a sustained, community-led coordinating organization for the PDE that mirrors the other Planetary Assessment or Analysis Groups, reports to the Planetary Science Advisory Committee, and meets regularly.
  NASA Chose to establish the position of the Chief Scientist of the PDE.
- ➢ Refine the full scope of the Planetary Data Ecosystem and build community consensus around it. The responsibilities, accountabilities, governance, and service levels for elements of the Ecosystem that are funded by NASA Planetary Science Division should be clearly defined.
  This is ongoing.
- ➢ The prioritized goals and scope of PDS need to be carefully and explicitly defined and clearly articulated to the community. The differing responsibilities and expectations of the data preservation mission versus distribution of usable data need to be clarified. PDS should not be given unfunded mandates.
  This is ongoing.

Data Preservation Needs High Priority Recommendations

Group 2: Address Data Preservation Needs

- Establish an archive for planetary radar data either within the PDS Small Bodies Node or separately. Time is of the essence to prevent irretrievable data loss.
  This is ongoing.
- Establish a requirement for the preservation of mission-supported laboratory analyses of returned sample material. Require data preservation with appropriate metadata in an approved archive or repository for data produced by laboratory analysis of returned samples supported by ROSES Data Analysis Programs.
  This is ongoing.
- Establish a carefully crafted strategy to identify and prioritize the data preservation needs of the planetary science community that are not currently being addressed.
- Consider ways of archiving outside of the PDS that are amenable to creating FAIR and standards-based archives of these growing data sets.
  These last two are items for which I am looking for Community input.

Barriers to Use & Development High Priority Recommendations

Group 3: Address Barriers to Use and Development - 1

- Include early funding for mission data acquisition, processing, and archiving of data and foundational data products so that they are planned well in advance of data acquisition.
  Looking for Input from AGs
- Training and outreach
  - ➢ Develop outreach to user communities within the Planetary Data Ecosystem, assess user needs, and develop focused educational and documentation materials that meet highest-priority needs.
  - ➢ Provide regular, accessible, and effective training programs for researchers, data producers, mission specialists, and others who need to archive with the PDS. Address data preparation from the perspective of reusability and interoperability, such as the Earth Science Data Systems Working Group (ESDSWG) Data Product Development Guide (DPDG) for Data Producers.
  - ➢ Expand opportunities for intermediate to advanced technical training in topics related to accessing, using, and processing planetary data.

Barriers to Use & Development High Priority Recommendations

Group 3: Address Barriers to Use and Development - 2

- Support the delivery of higher-level and analysis-ready data products in well-documented and broadly used protocols and formats.
- Broaden support across the Ecosystem for a wider variety of data and information formats, such as engineering data; data models; sound data; and physical collections.
- Expand intra- and inter-agency efforts to ensure that best practices, lessons learned, and appropriate technologies are shared and implemented across Planetary Data Ecosystem elements.
- The Planetary Data Ecosystem should regularly assess the Findability, Accessibility, Interoperability, and Reusability (FAIR) of data across each PDE element for machine-actionable access to data.

Barriers to Use & Development High Priority Recommendations

Summary:

- The concept of the PDE is meant to acknowledge that we have a wide diversity of people, data, data types, ways to store and preserve data, ways to access data, etc.
- Along with the richness of the wide diversity in the PDE comes a challenge of making sure everyone has access to these data… in perpetuity.
- I am working to lay the foundation for future Chief Scientists of the PDE to help NASA address this fundamental challenge.
- All of this needs your help.

Status Update on SPD-41: Scientific Information Policy

Open Source Science Initiative

Steve Crawford
Science Data Officer
NASA Science Mission Directorate

June 23, 2022

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

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	
1.1	Develop and implement a **consistent open data and software policy** tailored for SMD	**2.1**	Establish **standardized approaches for all new missions** and sponsored research that encourage the adoption of advanced techniques	**3.1**	Develop **community of practice and standards group**
1.2	Upgrade capabilities at existing archives to **support machine readable data access using open formats and data services**	**2.2**	Integrate investment decisions in High-End Computing with the strategic needs of the research communities	**3.2**	Partner with a**cademic, commercial, governmental and international organizations**
1.3	Develop and implement a SMD data catalog to support discovery and access to complex scientific data across divisions	**2.3**	Invest in capabilities to use commercial cloud environments for open science	**3.3**	Promote opportunities for continuous learning as the field evolves through collaboration
1.4	Increase transparency into how science data are being used through a free and open unified journal server	**2.4**	Invest in the tools and training necessary to enable breakthrough science through application of AI/ML		

SPD-41: Scientific Information Policy

SPD-41 was released in August 2021.

SPD-41 brings together existing NASA and Federal guidance.

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

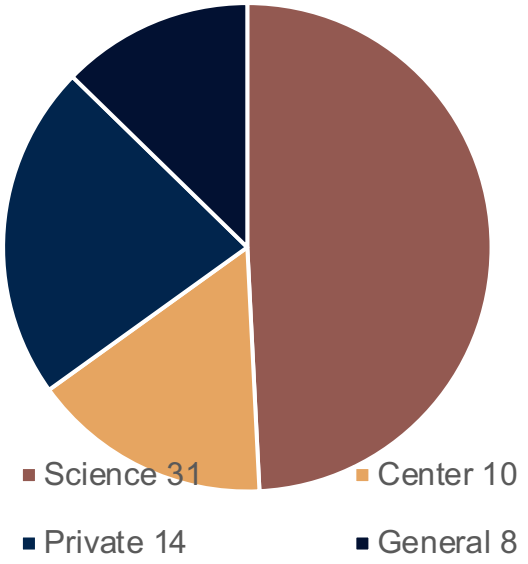
SPD-41a was released in November with proposed additions.
An RFI was released to the community and closed on **March 4, 2022.**

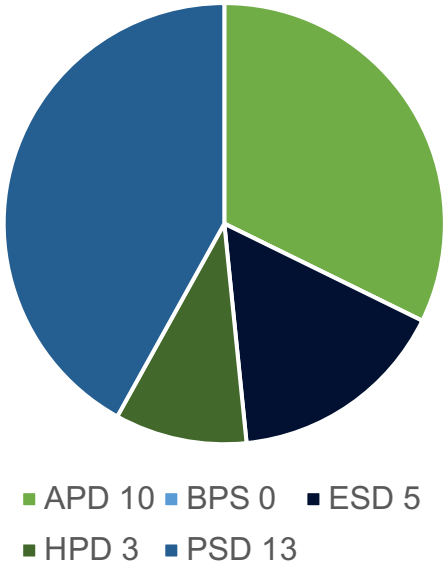
Summary on RFI responses from SPD-41

The RFI asked the following two questions:

- 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?

63 responses submitted to the RFI.





Overall, the responses were very supportive of Open Science and the goals of SPD-41.

There were concerns with the details of SPD-41 especially in software, support, and providing further guidance.

Summary of submitted information

- 125 suggested updates to SPD-41a submitted
  - 17 data-related suggestions
  - 39 software-related suggestions
- 87 areas to provide further guidance
  - 8 data-related
  - 31 software-related
  - 7 compliance
- 133 actions suggested to support implementation of SPD-41 including training, funding, compliance, and career incentives

Not all duplicates have been removed from this list.

Revision of SPD-41a

- SMD is currently in the process of revising SPD-41a based on the feedback received. This includes further clarification on scientific utility, software, conference material, and other areas.

- Each division will be provided an information policy providing further guidance on these specific areas for their communities.

- As a reminder, SPD-41a applies to future solicitations. Current grants or missions should adopt the policy consistent with available resources.

SPD-41 Schedule

Review

- June/July 2022: Revise SPD-41a based on the feedback received from the RFI.

Adopt

- August 2022: Adoption of SPD-41a

Support

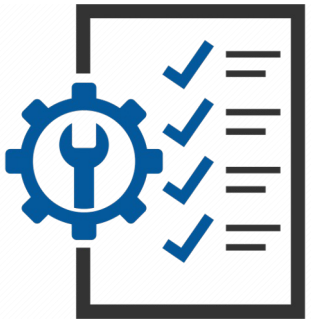
- Summer 2022: Transform to Open Science to include material relevant to SPD-41 in the core modules.
- Fall 2022: Host events to provide feedback on SPD-41a and RFI process.
- 2023: Training events as part of the Year of Open Science to include material on SPD-41.

Implement

- February 2023: Each division to release their updated data policies prior to ROSES23.
- January 2025: Compliance requirements will not begin before Jan 2025.

Open-Source Science Initiative

Unlocking the full potential of a more equitable, impactful, efficient, scientific future



Policy development, education, compliance tools
Updating NASA policies on scientific information to better enable the activation of open science

Core Services for Science Discovery
Developing core data and computing services to enable open science

ROSES Elements
Supporting open-source software, tools, frameworks, libraries, platforms, and training with over $5 million dollars in grants

Community Building & Partnerships - Transform to Open Science (TOPS)
Accelerating adoption of open science

CC BY

Leading the Path to Open-Source Science

Transform to Open Science (TOPS) is a $40 million*
5-year NASA Science Mission Directorate mission

Objectives:

- ★ Increase understanding & adoption of open science.
- ★ Accelerate major scientific discoveries.
- ★ Broaden participation by historically underrepresented communities.

2023 → 2024 → 2025 → 2026 → 2027

Year of Open Science 2023

Goals for 2027:

- ★ 20K earn Open Science Badge
- ★ 5+ major discoveries
- ★ Increase participation of underrepresented groups by 2x

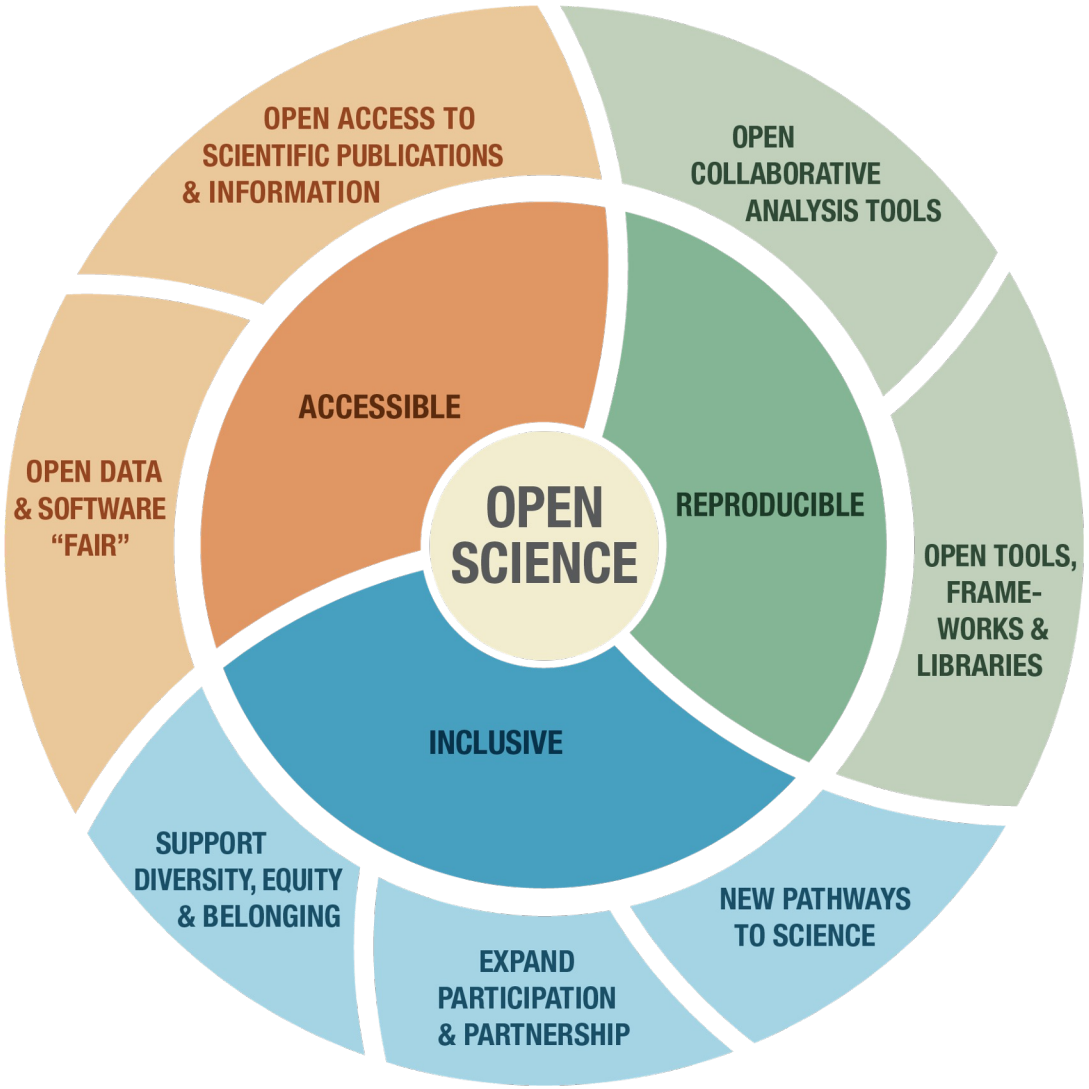
*pending appropriations

CC BY

BACKUP

What is Open Science?

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.



Science should be…

Transparent
scientific process and results should be visible, accessible, and understandable

Accessible
data, tools, software, documentation, and publications should be accessible to all (FAIR)

Inclusive
process and participants should welcome participation by and collaboration with diverse people and organizations

Reproducible
reproducible by members of the community

Open-Source Science is NASA's method to put O into practice.

- **Open** the entirety of the scientific process, ***from start to finish***
- **Broaden** community involvement in the scientific process
- **Increase** accessibility of data, software, & publications
- **Facilitate** inclusion, transparency, and reproducibility of science

SPD-41 Schedule to date

Timeline to Date:

- ✅ November 2020: Initial work on the policy started.
- ✅ August 2021: SPD-41 was adopted and released.
- ✅ November 2021: RFI on SPD-41a was released.
- ✅ March 2022: RFI closed and SMD received 61 submissions.
- ✅ May 2022: Submissions have been reviewed by OSSIC.
- ✅ June 7, 2022: OSSIC has reviewed the updated draft.

What is the current policy in SPD-41?

Data

Scientific data shall be made publicly available with a clear, open, and accessible data license no later than the publication of the research.

Mission data shall be openly available with no period of exclusive access.

Software

Research software should be publicly available no later than the publication of the research and assigned a permissive software license.

Publications

Manuscripts versions of as-accepted manuscripts shall be deposited in a NASA repository and made publicly available within 12-months.

Mission publications shall additionally be made publicly available at the time of their publication.

What are the new proposed changes in the SPD-41a draft?

Data

Scientific data should be FAIR and shall be made publicly available with a clear, open, and accessible data license no later than the publication of the research, and be citable.

Mission data shall be openly available with no period of exclusive access.

Software

Research software shall be publicly available no later than the publication of the research, assigned a permissive software license, and be citable.

Mission software shall additionally be developed openly in a publicly accessible, version-controlled platform that allows for contributions and engagement from the community.

Publications

Manuscripts versions of as-accepted manuscripts shall be deposited in a NASA repository and made publicly available within 12-months. Publishing as open access is supported and posting preprints is encouraged.

Mission publications shall additionally be made publicly available at the time of their publication.

Science workshops and meetings shall be open to broad participation and documented in public repositories.

Open science activities will be considered in reviews of proposals.