Planetary Data Ecosystem Independent Review

Presented by Melissa McGrath
Chair, PDE IRB

- NASA Planetary Science Division chartered the Planetary Data Ecosystem (PDE) Independent Review Board (IRB) in October 2020.
- The IRB's charter directed the team to:
  - define the full environment
  - identify missing or overly redundant elements
  - provide findings and prioritized recommendations
- Deliverables were a final report (submitted 15 April 2021) and a presentation to the PSD director (20 April 2021).

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

PLANETARY DATA ECOSYSTEM INDEPENDENT REVIEW

What is the Planetary Data Ecosystem?

The PDE IRB Charter contained a working definition:

> *"the ad-hoc, connected framework of activities and products that are built upon and support the data collected by planetary space missions and research programs which are primarily NASA-funded."*

and included a list of PDE Knowledge Elements:

ADS AMMOS AstroMat Autoplot

DAPs IPDA JMars HORIZONS

JSC Curation online catalogs MAPSIT

NASA Github NASA Planetary Github

NSSDCA PDART PDS

Planetary Geologic Mapping

Planetary Photojournal

PSIDA Quickmap RPIFs

The IRB's scope was very broad:

- All planetary sciences data (missions, telescope, lab experiments, in situ experiments, etc.)
- The full range of tools used for search and discovery, data analyses, data reduction pipelines, modeling and simulation tools
- Other software or firmware tools used by researchers to locate, calibrate, manipulate, and analyze these data

Archives, repositories, browse tools, software, organizations, physical facilities, funding programs…

The IRB Team

NASA Review Manager
Becky McCauley Rench
NASA Deputy Review Manager
Meagan Thompson

CTS Task Manager: Autumn Manecke
NRESS Support: Susan Borden
Latessa Tuck
Jason Silveira

Chair
Melissa McGrath

Subcommittees

Archiving
Co-Chair Moses Milazzo

Executive Secretary:
Sarah Black
NASA Ex Officio Members
Meagan Thompson
David Smith

Searching
Co-Chair Emily Lakdawalla

Executive Secretary:
Katya Gilbo
NASA Ex Officio Member
Lindsay Hays

Utilization
Co-Chair Sebastien Besse

Executive Secretary:
Don Hood
NASA Ex Officio Member
Sarah Noble

Mining & Automation
Co-Chair Bruce Wilson

Executive Secretary:
Hannah Dattilo
NASA Ex Officio Member
Megan Ansdell
KC Hansen

Inter-Relational
Co-Chair Caroline Coward

Executive Secretary:
Blake Mendoza
NASA Ex Officio Members
Steven Crawford
Lucas Paganini

Members

Kate Crombie
Martha Maiden
Joe Masiero
Flora Paganelli
Corrine Rojas
Priyanka Sharma
Marshall Tabetah

Julie Castillo-Rogez
Vandana Desai
Mary Ann Esfandiari
Laszlo Kestay
Matthew Miller
Lynn Neakrase

Alyssa Bailey
Ross Beyer
Henry Hsieh
David Mayer
Chase Million
Danielle Wyrick

Abigail Azari
Benjamin Burnett
Amitabha Ghosh
Jason Laura
Jennifer Shin
O. James Tucker
Alexandria Ware

Reta Beebe
Shawn Brooks
Stephane Erard
Robin Fergason
Patricia Lawton
Xiaogang Ma

Review Process

- Meetings from October 2020 through March 20 included:
  - ~60 Subcommittee meetings, many with invited speakers
  - Four full IRB meetings (November, December, January, February)
  - Appendix D of the final report lists all meetings, topics and speakers
- Full IRB meetings were open to the public, advertised via the DPS, LPI, AGU, and Planetary Society newsletters and via the science.nasa.gov/researchers/science-data website. Meeting recordings and minutes are also available at this website.
- The IRB maintained an open email address for input which was advertised in meeting announcements and during full IRB meetings
- NASA issued a Request for Information (RFI) on 17 August 2020
  - 28 Responses were received
  - List of Responses by title, with disposition for each, is included as Appendix E in the Final Report

Planetary Data Ecosystem Independent Review

The final 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

Core Values

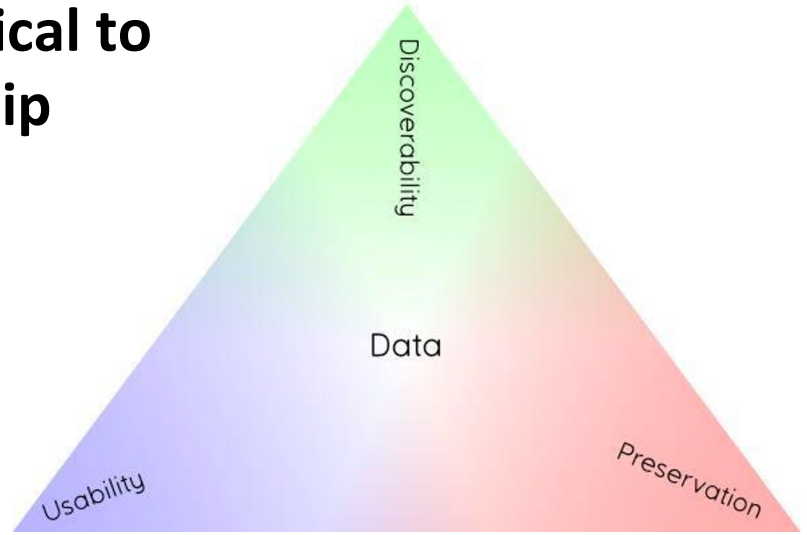
The IRB developed a set of core values

- **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.

Data Preservation, Discovery, and Usability

During the course of its review, the IRB repeatedly returned to the intricately intertwined concepts of data preservation, data discoverability, and data usability.

All three concepts are critical to successful data stewardship



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

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
- 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

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.
- **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**.
- 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.

Data Preservation Needs

Address Data Preservation needs

Plain language findings and recommendations summary

- Many critical data preservation needs are not being met. [The final report gives numerous examples.]
- A carefully crafted strategy is needed to establish priorities.
- Several data preservation needs are so urgent they should be addressed before an overall strategy is completed.

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.
- 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.
- 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.

Barriers to Use and Development

Address Barriers to Use and Development

Plain language findings and recommendations summary

- Many Planetary Data Ecosystem elements are designed for an expert group of users but should serve a broader user base.
- Many planetary data sets are difficult to use without extensive effort to convert them into formats compatible with modern scientific computing software, and to reduce low-level data to physical quantities of interest.
- ML/AI/AA and ARD should be utilized more effectively.
- As open as possible, as closed as necessary. Policy consistency and interoperability is desirable, at least across Planetary Science, preferably across SMD.

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.
- 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 and imaging data; and physical collections attached to planetary missions.
- 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**.

Toward an Ideal State

Pathway Toward an Ideal State

Prioritizing the many recommendations was the first step in forging a pathway toward an ideal state for the PDE.

Toward this end, the IRB also highlighted a particular case study, Small Body Science, as a powerful example that captures many of the key issues highlighted in the final report.

Addressing the challenges of Small Body Science provides one metric to measure our progress along the pathway toward **the ideal state for** the **Planetary Data Ecosystem**.

Additional case studies such as this, solicited from the broad user community, would be beneficial in assessing progress.

Toward an Ideal State

Small Body Science – a useful case study

The elements of the small-body ecosystem **are extremely diverse and present numerous challenges, some of which are shared with other planetary science fields, but several are unique.**

Unique challenges

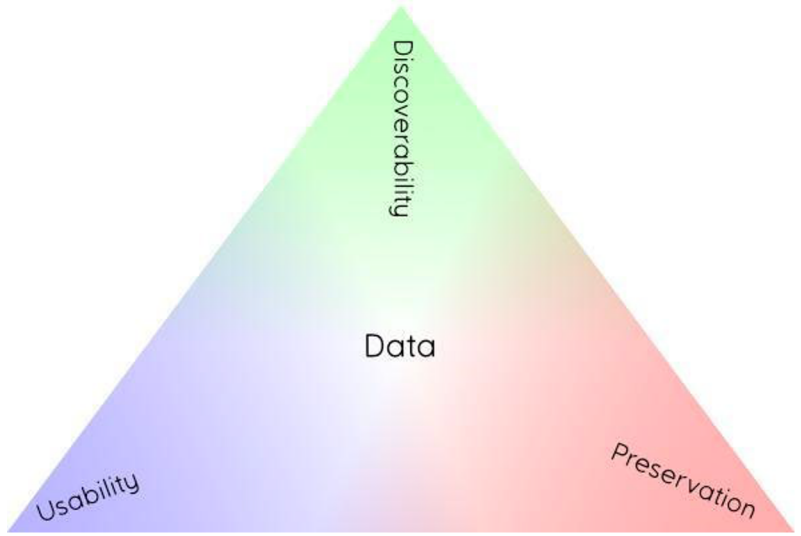
- The non-sidereal motion of small solar system objects means that useful data can be obtained serendipitously in observations of non-small-body targets. Systematically identifying and cataloging such serendipitous observations is difficult.
- Geometric and orbital circumstances of small-body observations are often essential to their interpretation but are not typically readily available to data users
- Given the significance of past, current, and upcoming wide-field surveys (e.g., Pan-STARRS, NEOWISE, ZTF, LSST, and many others) for small body science, techniques for analyzing large data sets are of increasing importance in this field.
- Time evolution studies also comprise an important aspect of small body science and require data covering time baselines that are as long as possible. Relevant data may be stored on older, less accessible media like photographic plates or magnetic tapes.

Toward an Ideal State

- Many facilities lack public archives, although their data could be critical in defining targets of space missions. Additionally, access to calibrated data (and calibration information) is challenging and hamper the scientific usage of those observations.
- Users need to seamlessly access various metadata (orbital parameters, observational geometry, etc…) that are often stored/archived/available at various locations (JPL's small body database, Minor Planet Center, PDS Small body node, etc.).
- Encouraging these metadata to be published would expand the reach of small body search tools, create opportunities for new collaborations, and lead to new discoveries.
- There is an important need to enhance integration among the various NASA tools, together with tools located at external institutions (e.g, CADC, ESA, ESO, etc…). This integration will strongly improve discoverability, and serendipitous detections.

Conclusion

The data gathered by the planetary sciences community is humanity's treasure. Along with NASA, and all elements of the Planetary Data Ecosystem, it is our responsibility to preserve and ensure its present and future usability.



Becky McCauley Rench
Program Scientist, PDS

Presentation to PAC on PDE IRB Implementation Status
June 14, 2021

Progress in Responding to PDE IRB Recommendations

- Develop the Planetary Data Ecosystem
  - CURRENT ACTION: PSD working group is identifying appropriate path to bring together a community-led group as identified by the PDE IRB. The first step will be for the community-led group to assist with further development of the PDE concept.
  - CURRENT ACTION: PSD will develop a webpage on science.nasa.gov dedicated to providing a centralized location for content about planetary data, including the current definition of the Ecosystem, its elements, and information on how PSD is addressing the PDE IRB recommendations.
- Address Data Preservation Needs
  - CURRENT ACTION: PSD currently supports radar data analysis, publication, and archiving of Arecibo data at the PDS SBN. PSD has initiated meetings between the SBN and the Arecibo, JPL, and Goldstone radar groups to coordinate formats and processes among their substantial radar data archiving efforts.
  - CURRENT ACTION: PSD is actively working preservation of mission-supported laboratory analyses of returned sample material for the OSIRIS-Rex mission. PSD will use this specific effort to identify and possibly address the broader needs of PSD laboratory sample data curation.
- Address Barriers to Data Use and Development
  - CURRENT ACTION: PSD working group met on June 9 to discuss next steps in this area.
- Other Recommendations
  - CURRENT ACTION: Specific response from NASA addressing ALL recommendations to include current status, anticipated timeline to address (if applicable), and potential future plans.

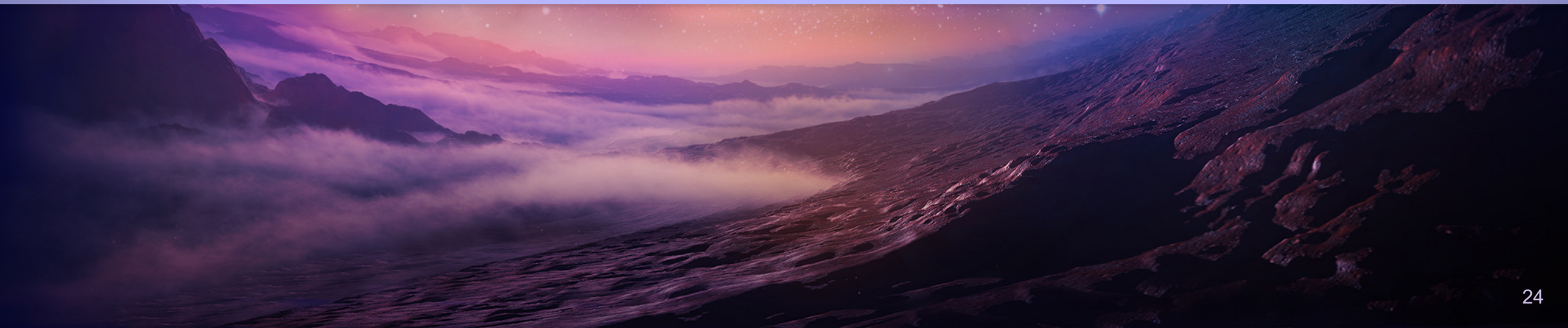
Schedule

- April 22, 2021 - NASA received the PDE IRB report
- May-June 2021 – PSD leadership and working group met to discuss first steps in implementation plan
- July 2021 – PDS Discipline Node Programmatic Review, which was structured to address some PDE IRB recommendations
- September 2021 – PSD Data webpage on science.nasa.gov will be available
- **Future Plans:**
  - End of 2021 - Specific response from NASA addressing ALL recommendations to include current status, anticipated timeline to address (if applicable), and potential future plans
  - Early 2022 – PDS Support Node Programmatic Review will continue addressing relevant recommendations

Accountability Plan

1) Provide regular updates on addressing the PDE IRB recommendations using the new PSD data website
2) Report to the internal SMD Strategic Data Management Working Group on progress
3) Provide status updates to the PAC as part of regular reports from PSD Director and/or R&A Lead, with specific details when milestones are achieved

BACKUP: MORE DETAIL



Develop the Planetary Data Ecosystem

CURRENT ACTION: PSD working group is identifying appropriate path to bring together a community-led group as identified by the PDE IRB. The first step will be for the community-led group to assist with further development of the PDE concept. (**R1**).

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)

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.

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.

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.

Develop the Planetary Data Ecosystem (cont.)

CURRENT ACTION: PSD will develop a webpage on science.nasa.gov dedicated to providing a centralized location for content about planetary data, including the current definition of the Ecosystem, its elements, and information on how PSD is addressing the PDE IRB recommendations. The plan is for this website to be publicly available in Fall 2021 and be similar to https://science.nasa.gov/heliophysics/heliophysics-data.

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.

Address Data Preservation Needs

CURRENT ACTION: PSD currently supports radar data analysis, publication, and archiving of Arecibo data at the PDS SBN. PSD has initiated meetings between the SBN and the Arecibo, JPL, and Goldstone radar groups to coordinate formats and processes among their substantial radar data archiving efforts.

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 (cont.)

CURRENT ACTION: PSD is actively working preservation of mission-supported laboratory analyses of returned sample material for the OSIRIS-Rex mission. PSD will use this specific effort to identify and possibly address the broader needs of PSD laboratory sample data curation.

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.)

CURRENT ACTION: PSD working group is identifying possible actions to be taken at NASA HQ in this area.

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.

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

CURRENT ACTION: PSD working group met on June 9 to discuss next steps in this area.

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.

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.

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.

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.

END SLIDE