Final Report

of the

Planetary Data Ecosystem Independent Review Board

April 2021

Table of Contents

Executive Summary

The Planetary Data Ecosystem (PDE) Independent Review Board (IRB) was chartered by NASA in the fall of 2020 to conduct a wholistic review of the Ecosystem with the goals of "defining the full environment, identifying missing or overly redundant elements, and providing findings and prioritized recommendations." The IRB was charged to complete its review by the end of March 2021. This report and the associated briefing package to NASA constitute the deliverables from the IRB's work.

Like the Ecosystem, the scope of the IRB's effort was very broad. The high-level goal of this IRB is to help NASA to develop a seamlessly integrated Planetary Data Ecosystem that improves the planetary science community's access to, and use of, high-quality data. Five months of gathering and thoroughly discussing input has culminated in the 67 Findings and 65 Recommendations presented below. These fall into three broad categories, each of which is discussed in one or more Sections of the report: the continued strategic development of the overall Ecosystem (Sections 2 and 3); barriers to data preservation (Sections 3 and 4); and barriers to access, usability and development (Sections 5 and 6). Our Recommendations have been prioritized, however, due to the large number of Recommendations, we present only the highest-priority Recommendations in each of these three broad categories.

Foster the strategic development of the Planetary Data Ecosystem to build upon NASA's investment in the Planetary Data System (PDS) and address the presently unmet usability and data archival needs of the planetary science community. This development includes ensuring that a standing assessment or analysis group exists for the Ecosystem. Such a group is needed to help establish and maintain effective, on-going, multi-way communication between NASA, the broad planetary sciences community, and the elements of the Ecosystem. The group will help ensure that the concept of the Ecosystem is well communicated to the community, and that the needs of the community are clearly communicated to NASA. It is imperative that this group be broadly representative of all aspects of the user and developer community. This group is essential to help NASA continue to refine the full scope of the Ecosystem, to take the recommendations of this IRB and develop a long-term strategy, and to provide guidance as circumstances and knowledge evolve. The broader Ecosystem and stewardship of planetary data are discussed in Sections 2 and 3.

Address barriers to data preservation, including two time-critical needs, planetary radar data and returned sample data. Although the overall intent of this group of Recommendations is to prioritize development of a community strategy, these two data preservation needs are considered more critical and should take precedence over the many additional preservation needs expressed by the community. The additional data preservation needs are discussed in detail in Section 4.

Address barriers to the use of planetary data and future development around these data. Multiple structural and technical factors limit the degree to which planetary data is Findable, Accessible, Interoperable, and Reusable (FAIR). Many of these findings echo those from previous assessment reports and most affect the broad range of Ecosystem data users. These barriers have a particularly strong effect limiting the usability of planetary data for some of the most data-intensive efforts, including those working in machine learning, artificial intelligence, and other advanced analytics methods. These barriers significantly limit the scientific return on investment for the resources spent on acquiring planetary data. These barriers to access, usability and development are discussed in detail in Sections 5 and 6.

We conclude in Section 7 with a particularly useful case study that well illustrates the need to address Recommendations in all three of the categories outlined above as we follow a pathway to a more ideal future state for the Planetary Data Ecosystem.

Plain Language Summary

In October 2020, NASA chartered an Independent Review Board (IRB) to evaluate its Planetary Data Ecosystem (PDE). The idea of a Planetary Data Ecosystem is a new one; NASA defined it to include the user community and the organizations and facilities involved in planning, obtaining, analyzing, preserving, and sharing data from planetary space missions. It is not centrally planned but has grown from the activities of people working in it.

NASA charged the IRB with evaluating how well the PDE is serving the community who use (or would like to use) planetary data, and with providing prioritized recommendations to NASA for how to take actions to improve the PDE. The IRB membership included experts in such fields as earth science, planetary and atmospheric science, data and information science, archives, and many others. Most members were affiliated with organizations outside of NASA, including a few outside the United States. The IRB charter established five subcommittees (Archiving, Searching, Utilization, Mining & Automation, and Inter-relational) and tasked each subcommittee with examining particular aspects of the ecosystem.

Among IRB recommendations are the establishment of a standing assessment or analysis group to establish and maintain effective communication between the broad planetary sciences community and NASA's PDE, urgent attention to preservation of planetary radar data and preservation related to returned samples, consideration of additional ways of archiving outside of the standard and well-established data repositories, and other suggestions to fulfill needs not currently being met. This report is one of two deliverables mandated in the charter. The other was a live presentation to NASA, which took place on 20 April 2021.

1 Introduction

1.1. Review Process

The NASA Science Mission Directorate Planetary Science Division (PSD) chartered a board in the fall of 2020 to perform an independent review of its Planetary Data Ecosystem (PDE), a newly defined entity that encompasses the broad spectrum of activities that take PSD data from creation to dissemination.

The Independent Review Board (IRB) was tasked with providing findings and prioritized, actionable recommendations from which an optimized strategy can be developed for the PDE. By charter, the IRB structure was directed to include five subcommittees based on functional aspects of the Ecosystem: Archiving, Searching, Utilization, Mining & Automation, and Inter-relational. Each subcommittee subsequently defined its own charter. The overarching IRB charter and each of the subcommittee charters are included in Appendix A.

The IRB membership was initially solicited by NASA and vetted through Cornell Technical Services (CTS). The selected Chair then led selection of the subcommittee Co-Chairs and the combined leadership subsequently selected the subcommittee membership from among a large pool of candidates assembled by NASA and CTS. A set of guidelines was used to ensure broad representation from key areas as well as a range of geographically and internationally dispersed science community members. Members included geologists, planetary and atmospheric scientists, data managers, archivists, community leaders who direct the daily effort of managing science data processing from both the Earth Science and Astrophysical communities, as well as those communicating and representing NASA science to the general public. Each subcommittee was assigned one or two NASA Ex-Officio members and each subcommittee Co-Chair selected an Executive Secretary.

The IRB utilized a Steering Group to help organize and direct the day-to-day activities, as well as to prepare for the full IRB meetings. The Steering Group consisted of the NASA Review Manager; the CTS Task Manager; the Chair and Co-Chairs; the NASA Ex-Officio members; and the Executive Secretaries. The Steering Group met on an as-needed basis. An IRB executive group, composed of the NASA Review Manager and deputy Review Manager, the CTS Task Manager, and the Chair met on a weekly basis during the review period. The work of the IRB was supported throughout its activities by both CTS and NASA Research and Educational Support Services (NRESS). The structure and membership of the IRB is included in Appendix B.

A Steering Group organizational meeting was held on 19 October 2020 to finalize the IRB membership. The full IRB kick-off meeting was held on 12 November 2020. Between November 2020 and March 2021, three full IRB meetings were conducted spaced approximately monthly in December

2020, January 2021 and February 2021. Due to the worldwide coronavirus pandemic, all meetings were held virtually utilizing NASA-supported Google tools. The full IRB meetings were open to the public. Meetings were announced via a NASA website, and via community newsletters including the Planetary Exploration Newsletter (PEN), the Division for Planetary Sciences (DPS) newsletter, the Lunar and Planetary Institute (LPI) newsletter, and the American Geophysical Union (AGU) Planetary Section newsletter.

Subcommittees generally met once per week. Discussions were lively, focused and resulted in the contextual background, findings and recommendations within this report. In addition to the regular meetings, subcommittees sought external voices, scheduled tutorials, and solicited external materials or briefings from a range of areas. All of these contributed to the overall knowledge and understanding of the IRB members. Details about the approach used by each subcommittee are included in Appendix C. Appendix D contains a list of the full IRB, Steering Group and subcommittee meetings held by the board throughout the review. The NASA website mentioned above includes IRB meeting recordings and minutes.

NASA also solicited community input via a Request for Information (RFI). The RFI officially closed on 9 November 2020 with a total of 28 responses. Appendix E lists all RFIs by title and number. A disposition for each RFI is also included in the table in Appendix E.

Another technique used to solicit community input throughout the review period was a publicly available email address (PDEIRB.Input@nress.org), which was advertised in meeting announcements and presentations.

To support its work, the IRB established a Curated Document Library, the contents of which are included in the reference material detailed in Appendix F. A list of acronyms is included as Appendix G, and a Glossary is included as Appendix H.

The IRB deliverables are outlined in the Charter (Appendix A) and summarized here:

1. A non-consensus report with observations, findings, concerns, and prioritized, actionable recommendations to the PSD Director.
2. A presentation to the PSD Director and other NASA stakeholders summarizing the review results.

The report is structured as follows:

- Section 2 discusses the overall concept of the Planetary Data Ecosystem.
- Section 3 addresses the topic of NASA stewardship of planetary data.
- Section 4 concerns some of the needs not being met in the current PDE.
- Section 5 identifies barriers to access and usability of planetary data.
- Section 6 articulates barriers to development.

- Section 7 describes a potential pathway to an ideal future state of the Ecosystem using a prioritized set of recommendations.
- Extensive supporting materials are provided in a set of Appendices.

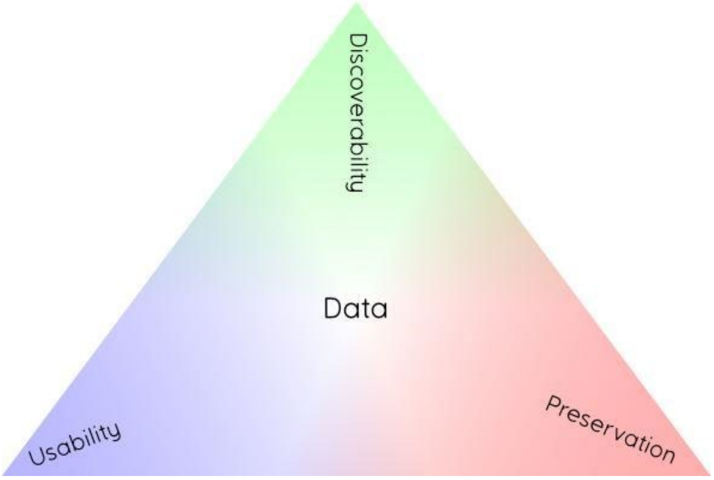
Non-consensus recommendations are clearly indicated, and a brief explanation of the non-concurrence is also provided.

This final report was reviewed by NASA in advance of its release to ensure that its scope meets the intent of the IRB's charter, and that this report's findings and recommendations are clear.

1.2. Data Preservation, Discovery, and Usability

The scope of the Planetary Data Ecosystem is broad. It encompasses not only the full landscape of science data gathered by researchers using land- and space- based instruments, telescopes, lab and in-situ experiments, observation missions, etc., but also the full range of tools used for search and discovery, data analyses, data reduction pipelines, modeling and simulation tools, and other software or firmware tools used by researchers to locate, calibrate, manipulate, and analyze these data.

The IRB repeatedly returned to the inextricably interconnected concepts of data preservation, data discoverability and data usability. Without preservation, data are irrevocably lost, with no hope of ever gleaning information from the data again. However, and *equally critical*, data gathered and preserved without regard to future discoverability and usability in mind renders the data somewhat analogous to Crown Jewels: well and carefully preserved so that we can admire their beauty, but never again available for use. This concept of the critical intertwining of the three aspects of data stewardship is illustrated in the diagram below. Not only is each corner (color) of the triangle required for seamless data stewardship, but the blending among the colors anchored in the corners of the triangle is also required if we are to make the most of the data we have spent enormous resources gathering.



1.3. IRB Core Values

During the course of its activities the IRB developed a set of core values that encapsulate our overarching vision:

- **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.

It is with this shared spirit that we offer the findings and recommendations in this report.

2 The Planetary Data Ecosystem Concept

2.1. Introduction

NASA provided the IRB with an initial definition for the Planetary Data Ecosystem (PDE or “the Ecosystem”): “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 primarily are NASA funded.” Although members of the planetary science community may think of the Planetary Data Ecosystem as the place where NASA keeps its data, the concept actually recognizes a much broader landscape that includes the discovery, access, and use of planetary data, not only preservation. In fact, the Ecosystem extends beyond NASA, encompassing a global myriad of elements that focus on information relevant to planetary science. The concept of a data ecosystem arises from work on information ecology, which is a system of people, practices, technologies, and values in a local environment (Nardi and O’Day, 1999).

Before a spacecraft acquires digital data or returns samples, science teams plan their data acquisition and initial analysis. Engineers design, build, and test hardware. Software developers build the sensor models, processing pipelines, and targeting code to support data collection and initial processing. Once data and/or samples have been transmitted, investigators process, calibrate, reduce, summarize, mosaic, map, and otherwise analyze the data to produce higher-level data sets. Not all planetary data arrives from space; scientists work in laboratories and observatories, develop software and models, and study at analog field sites around the world. Investigators write abstracts, publish research, and speak at conferences. Media relations offices document all these activities, producing photos, videos, podcasts, press kits, press releases, and media events surrounding a mission’s operational activities and the resulting scientific discoveries. Archivists and historians preserve records in archives and libraries. Public engagement and education offices produce websites, outreach products, formal education curriculum, teacher training materials, planetarium shows, and museum exhibits. Museums, planetariums, and visitor centers produce their own exhibits, shows, and educational materials. Members of the public process data and participate in formal or informal citizen-science campaigns. NASA directly funds many of these activities, but not nearly all of them.

NASA communicated the Planetary Data Ecosystem concept to this Board as an initial list of “PDE Knowledge Areas” (see Appendix I) and as lists of elements to be examined by each chartered subcommittee (see Appendix A). This list was meant to be illustrative, not comprehensive. To investigate the current state of the Ecosystem, the IRB needed to develop a more thorough list of elements and a working concept of the information it includes.

2.1.1. Planetary Data Ecosystem Information

For the purposes of this review, the IRB included the following types of information in the working concept of the PDE:

- **Data returned from space missions and ground-based facilities** including observational data, telemetry and other engineering data, samples, and mission planning documents;
- **Data generated by research and analysis projects** including observational data analysis, theoretical research, laboratory results, and Earth analog site field tests;
- **Data generated by citizen scientists**, including participants in observation campaigns, contributors to collaborative citizen-science services, and space enthusiasts;
- **Standards** for planetary science data and metadata;
- **Software** including data processing pipelines, analysis tools, search and browse tools, display tools, and simulation tools;
- **Publications** including articles, books, conference abstracts, reports, posters, and presentations; and
- **Education and communication products** including value-added products from missions and facilities (websites, captioned photos, etc.), educational materials, recordings of outreach events, products generated for the media, and unpublished photo and video documentary material gathered for public engagement purposes.

2.1.2. Planetary Data Ecosystem Communities

The following are types of communities that the IRB considered to be participants in the Ecosystem:

- Personnel from NASA and other space agencies (themselves containing many different stakeholder groups);
- Mission and ground-based facility personnel;
- Science researchers, technology innovators, software developers, media professionals, historians, artists, and others who use planetary data in a professional capacity;
- Amateurs, enthusiasts, and hobbyists; and
- Educators, students at all age levels, and parents of students, in both formal and informal education environments.

2.1.3. Archives, Repositories, and Registries

Planetary Data Ecosystem elements are designed and utilized by distinct communities. Some serve the needs of multiple communities; others focus narrowly on use by only one. They vary widely in discoverability, accessibility, searchability, and usability. Table 2-1 provides some examples of the elements associated with each kind of information and example users, while Appendix I offers a longer (though still not comprehensive) list.

A major challenge in evaluating the current state of the PDE was a semantic one. There are many words used to describe places where information is preserved, but the words can mean different things to different people. In this document, we consider three major types of facilities that preserve and manage access to information: archives, repositories, and registries. Also, facilities can be physical (e.g. a curated sample collection), virtual (e.g. an online access point to planetary data), or both (e.g. a library that has digitized part of its material collection for remote access).

The term “archive,” in particular, has multiple distinct meanings. Although both digital and physical archives have the most rigorous standards for ingesting material into their permanent collections, they differ in accessibility to the public. Also, when a member of the planetary science community discusses “archiving” data, they might mean to say they are “contributing to an archive,” but they may also mean “depositing in a repository” or “listing in a registry.” These multiple meanings have led to confusion about how to satisfy federal mandates to share the results of federally funded research with the public (see, for instance, the text following Finding 54 about GitHub not being an archive).

The following are definitions of how the words “archive,” “repository,” and “registry” are used in this report.

2.1.3.1. Physical Archives

A physical archive is a collection of historically significant items that are selected, curated, organized, and preserved for long term discovery, access, and use. Archivists often create finding aids that catalog groups of related items to assist with access. Archivists also employ highly specific preservation methods to ensure that archival assets are available for use well into the future. In planetary science, archives of physical samples are important and have widely varying archival needs. Most physical archives now provide digital access to some part of their collections and are actively working to digitize physical items as funding permits. Examples of physical archives in the planetary sciences include:

- The former NASA Regional Planetary Image Facilities (RPIFs)
- The Astromaterials Acquisition and Curation Office at Johnson Space Center (JSC)
- Archives located at each of the NASA Centers and NASA Headquarters

2.1.3.2. Digital Archives

Digital archives have the following properties: they are independent, or managed by someone other than a major data provider; sustainable, managed for the long term (25 years at least); searchable; citable; preeminent, or considered by its user community as the “standard” archive for the subfield; and standardized, providing data products in standardized formats and file types. In addition, it is desirable for archives to be open-access and to provide peer review and documentation (user guides, calibration descriptions, etc.) for the data they hold. This definition comes from language in NASA’s

Planetary Data Archiving, Restoration and Tools (PDART) announcements of opportunity. Example digital archives that contain data relevant to planetary science include:

- The nodes of the Planetary Data System (PDS)
- The Center for Near Earth Objects Studies (CNEOS)
- The High-Resolution TRANsmission molecular absorption database (HITRAN)

2.1.3.3. Repositories

Repositories, whether physical or digital, have less stringent standards for inclusion than archives, but are often curated to enable search, discovery, and reuse of their content. They may preserve, manage, and provide access to many types of materials in a variety of formats. Repositories vary widely in their standards and level of curation. A useful repository for the storage of science data must have sufficient control for digital material to be authentic, reliable, accessible, and usable on a continuing basis. Throughout this report, we will refer exclusively to digital repositories. Example repositories include:

- Astromaterials Data Repository (AstroMat)
- Institutional data repositories, e.g. those at the Smithsonian, APL, and LPI
- Scientific journals that provide data underlying peer-reviewed articles
- Various software repositories available on the internet (e.g., GitHub)
- NASA’s data.nasa.gov

2.1.3.4. Registries

Registries are catalogs. They do not hold data; Instead, they describe the location, access mode, and (to some extent) the content of tools, data sets and/or products. Registered products can include data files, label files, metadata and other schemas, dictionary definitions for objects and elements, services, and so on. Examining these products can give users insight into the structure, organization, and (to a lesser degree) the contents of related data sets. Registries are typically not curated, relying instead on data providers to submit and document content. However, archives sometimes host registries of data holdings at other locations that are within the scope of the archives (e.g. lists of data holdings for non-NASA missions from PDS archives of NASA-funded instruments on those missions). Currently, most registries are related to the medical and health sciences fields, but there are many others. Examples include:

- The U.S. EPA System of Registries
- The Open Metadata Registry
- Kwantu Data Registry
- The NIH National Institute of Environmental Health Sciences Sample Collection Registry

Table 2-1 (below and next page): Examples of the diversity of elements potentially contained in the Planetary Data Ecosystem. It is not an exhaustive list; the examples are intended to illustrate the wide variety of types of data, the large quantity of data locations, and diversity of data users and goals in the Planetary Data Ecosystem. A more detailed list of data repositories is available in Appendix I. The approach to generating a list of potential Ecosystem data users is described in Appendix C.

Type of Data	Location	Example User
Data/samples returned from space missions and ground-based facilities		
PSD-funded mission data (through Phase F)	PDS for raw, calibrated, and derived products under formal agreement; some mission websites (e.g., LROC)	A software developer building a new mission visualization tool
Mission planning documents (e.g., science plans, working group reports)	Science plans sometimes delivered to PDS; summary of planning processes sometimes published in peer-reviewed publications; other mission documents generally not archived	An early-career scientist who is trying to assemble their first mission proposal and wants to leverage what was done in the past.
Observations from large telescope facilities, including radar facilities	PDS; dedicated archives (e.g., IPAC, MAST)	An illustrator building an infographic showing sizes and shapes of near-Earth asteroids
Observations from small telescope facilities	PDS; MPC; CNEOS	An undergraduate researcher interested in lunar flashes
Data generated by research and analysis projects		
Mission data analysis	NASA Open Data and Software Portal; PDS; USGS publications (e.g., maps)	A newly selected participating scientist who needs to get up to speed on mission science results
Theoretical research	NASA Open Data and Software Portal; NASA STI repository; ADS	A scientist proposing a Discovery mission developing a science traceability matrix
Laboratory results	NASA Open Data and Software Portal; RELAB; NASA STI repository; ADS; PDS	An early career scientist proposing new laboratory work who wants to build on the state of the art
Analog field tests (including photos, notebooks, rock and water samples, geophysical measurements)	Mostly in researcher's desks/hard drives; password-protected websites; some in community databases; very small fraction in the PDS or PDS-equivalent repositories (e.g., USGS ScienceBase); AHED	A high school educator looking for planetary analogs for a geography class An engineer developing a drilling tool for a planetary lander seeking analog drillable materials
Standards		
Planetary coordinate systems	International Astronomical Union (website maintained by USGS with NASA funding)	A mission planner trying to locate a landing site on an asteroid
Planetary nomenclature	International Astronomical Union (website maintained by USGS with NASA funding)	A high school student wanting to know why all the craters on Venus have female names
Planetary body ephemerides	NASA/JPL Solar System Dynamics; HORIZONS	A scientist planning mission encounter scenarios for her mission.
Planetary body physical properties	NASA/JPL Solar System Dynamics; Minor Planets Catalog (MPC)	A postdoc who needs the value of Mercury's moment of inertia to calculate the size of its core

Type of Data	Location	Example User
Software		
Software for theoretical research (e.g., modeling)	NASA GitHub	A research scientist benchmarking their own software
Software pipelines for processing mission data	Researchers' desks and hard drives	A research scientist who would like to reprocess raw data to better understand their characteristics
Software for data manipulation and processing (e.g., mapping)	USGS-supported software tools (e.g. ISIS); JMARS; NASA Treks; commercial packages	An educator creating maps of planetary surfaces
Software for planetary body orbit predictions	NASA SSD; HORIZONS; Cosmographia	A journalist advertising objects that are observable with amateur telescopes every month
Publications		
Peer-reviewed papers	NASA STI repository	A congressional staffer creating a graphic of NASA-funded research results over the years
Conference abstracts	ADS	A scientist searching for preliminary results in a field of interest
Selected NASA Center internal newsletters and project reports	NASA Center archives	A historian writing a book about the challenges encountered during the development of a mission
Data generated by citizen scientists		
Data generated under NASA Citizen Science	Sometimes supplemental materials in peer-reviewed journal articles	Researcher investigating the timing of the formation of different ice features on Mars
Products based on NASA-funded data via private entities (e.g., Zooniverse)	Publications; conference abstracts	A citizen scientist who would like to become involved in the analysis of space data
Education and communication materials		
NASA press releases and press kits	NASA Mission websites; NASA Center archives	A science writer summarizing the top science discoveries from a mission
Captioned, processed, publicly released images	Websites of missions, instrument teams, NASA centers, individual researchers, etc.; photo galleries on Flickr, Facebook, Instagram, etc.	A 5th grade student researching a school project on a planet A planetary science professor searching for images for a class presentation
Unpublished media products (e.g. spacecraft assembly & testing photo & video)	NASA Center archives	A book publisher looking for photos of people of color at work in spacecraft assembly and testing
Educational products	Websites for NASA missions, NASA Education, or non-NASA (PI-led) mission websites	A 3rd grade teacher looking for a standards-based classroom activity for English learners

2.1.4. FAIR (Findable, Accessible, Interoperable, Reusable) Data Principles

The FAIR data principles enumerate and explain a set of qualities that enable the discovery and integration (both human and machine) of scientific data leading to knowledge sharing and innovation (Wilkinson et al., 2016). They include making data:

- **Findable:** The location and metadata of data sets are discoverable using human and machine tools, such as standard citation formats in the literature or web search engines, via unique, persistent identifiers.
- **Accessible:** Data sets and related metadata are able to be viewed, downloaded, and transferred via appropriate authentication or authorization methods.
- **Interoperable:** Data sets are easily integrated with other data, applications, and workflows through common metadata.
- **Reusable:** Data and related metadata are able to be re-analyzed and replicated in a variety of electronic environments.

FAIR data are often mentioned concurrently with the “Open” movement (i.e. open data, open science, open software, and so on). However, they are highly distinct: data do not have to be open-access to be FAIR. A repository may require permission, authentication, paid subscription, or some other perceived barrier to overcome before gaining access to any content. Similarly, open data may not follow the FAIR principles. A repository may be freely accessed, but may not be discovered by common web search engines, may have insurmountable barriers to retrieving data with the user interface, may be so stovepiped as to be inaccessible to other systems, and may not be usable in any other context.

FAIR data is also often judged based on the perspective of an interactive user navigating through web pages, for example. However, the FAIR principles emphasize machine-actionability. For data to be FAIR, computational systems must be able to find, access, interoperate, and reuse data with little to no human intervention. This is necessary given that humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.

The FAIR Principles apply to all types of data, including related analytical software, tools, and metadata.

2.2. Findings and Recommendations: The Planetary Data Ecosystem Concept

2.2.1. Establish the Planetary Data Ecosystem Concept

Finding 1 The IRB endorses NASA’s concept of a Planetary Data Ecosystem that includes not only data but also the communities that interact with it.

The foundation of the Ecosystem is the planetary sciences **community**: the people who produce, provide, and use data (mission personnel, research and analysis (R&A) scientists, lab technicians, theoreticians, sample collectors and curators, engineers, software developers, historians, documentarians, journalists, educators, citizen scientists, enthusiasts, artists, students, and others).

The Ecosystem encompasses a broad array of **information** related to planetary systems, including both raw data (whether it be returned from missions, observatories, analytical laboratories, fieldwork, theory, model development, etc.) and derivative products (higher-level data products, maps, simulations, models, books, articles, posters, lectures, videos, artworks, media communications, educational materials websites, etc.)

Finding 2 Adopting the Planetary Data Ecosystem concept could improve preservation, accessibility, and usability of planetary data.

Finding 3 The concept of the Planetary Data Ecosystem needs further clarification.

During the information-gathering phase of the conduct of this review, IRB members observed that the idea of the Planetary Data Ecosystem generated excitement and interest among other members of the planetary exploration community that the IRB contacted. The Ecosystem has potential to be a useful framework within which common science investigations could be shared, findings expanded, duplication reduced, and needs prioritized. It could enable different Ecosystem elements to focus on particular communities and subject domains while interoperating across a cooperative ecosystem; help to ensure that valuable information is preserved for posterity in a way that is discoverable and has a high degree of usability; and could foster innovation and experimentation even while missions are actively collecting data.

However, neither the working NASA definition of the Ecosystem, nor the list of elements developed by this Board, defines the scope of the Planetary Data Ecosystem clearly enough to determine if a specific element is within the scope of the PDE or not. Those who have not heard of the PDE concept often assume that it is the same as NASA's Planetary Data System (PDS). Although the PDS is a cornerstone of the PDE, it is only one component. This panel has developed a list of proposed PDE Elements (Appendix I), but this list is not comprehensive. Even among those who have heard of the concept, there are different interpretations. The definition provided by this panel is a starting point, but it is important that all of the stakeholders in the PDE have a common understanding of what the PDE is and how the different elements work together. Ultimately, NASA must develop a definition of the Planetary Data Ecosystem to differentiate the elements that fall within NASA's purview from those that fall outside it.

The Earth Observing System Data and Information System (EOSDIS) concepts of “core” and “community” may be a useful model. EOSDIS uses the concept of “core” to refer to those things directly funded by the Earth Science Data and Information System (ESDIS) Project. “Community” refers to those things funded by other programs (including other elements of the Earth Science

Data Systems program) that are part of the broader Earth science data ecosystem. Community systems and projects can take on higher risk and employ technologies that are in earlier stages of development. In order for the new technologies developed in the community to infuse into the core, they must be developed in a way that can interoperate with existing core technologies. By targeting community systems and projects for outreach and communication around interoperability, NASA can encourage technology infusion.

Recommendation 1 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.

Recommendation 2 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.

Recommendation 3 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.

2.2.2. Encourage Development of the Planetary Data Ecosystem as a Community

Finding 4 NASA’s Planetary Science Division is only one of many entities with roles within, and responsibilities to, the Planetary Data Ecosystem.

Several PDE elements are not funded by NASA’s Planetary Science Division, limiting Planetary Science Division control.

Finding 5 Insufficient communication and collaboration has resulted in inconsistent practices and standards that prevent interoperability across Planetary Data Ecosystem elements.

Even within NASA, PDE elements are governed by many divisions and offices that often do not collaborate in their stewardship of planetary science information. Inconsistencies exist in data management and archiving practices; data access modalities; security policies, practices, and procedures; and the types and formats of repositories that hold data. Taxonomies, standardized metadata, and controlled vocabularies are rarely shared across PDE elements. This is not surprising, given that the concept of treating these elements as an Ecosystem is new, but it remains an opportunity for improvement going forward.

Finding 6 There is, at present, no broadly representative community users' group or advisory body that coordinates and prioritizes tasks and recommendations among the Planetary Data Ecosystem entities, rendering effective collaboration difficult.

Finding 7 There is no mechanism for NASA to solicit input from the disparate Planetary Data Ecosystem communities, contributing to NASA's unilateral approach to data management.

Finding 8 There is no mechanism for transparent, widespread communication of plans, timelines, or developments of new or changed capabilities, either from NASA to other Planetary Data Ecosystem communities and elements, or among the elements of the Ecosystem.

Recommendation 4 **(Non-consensus)** 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.

The charter for this organization should include: regularly soliciting community input; refining the scope and expanse of the Ecosystem, including how the PDS fits within the Ecosystem; frequently communicating plans and progress to the Ecosystem's communities; and making findings and recommendations to the Planetary Science Advisory Committee for improvements in the Ecosystem. Care must be taken that the business of this organization be conducted in a way that is accessible for global participation. Transparency, communication, and coordination among all elements of the Ecosystem are key.

This organization should include a diverse membership from as representative a sample of the full Ecosystem as possible. This organization should ensure that all domains are specifically represented. Its role would include the identification of gaps in software tool capabilities and analytics related to the Ecosystem, enumerating the basic features that would be required to fill those gaps, and ranking of the gaps by criticality of need.

Without the existence of such a sustained, coordinating organization, the long-term future of the preservation, availability, accessibility, and usability of NASA's planetary data sets will continue to remain uncertain. This Report is only the latest in a 39-year history of documents that advise NASA on how to improve the preservation and accessibility of the results of its funded missions and research (see, e.g. NRC 1982; PDS Roadmap Study 2017; and section 3.1). The IRB observed that many findings and recommendations could be tracked, in similar forms, through this history of reports, indicating that many problems identified by review panels have remained unaddressed. Some, but by no means all, of these parallels are pointed out in this report. The consistency over time of some findings and recommendations demonstrates the need for a sustained advisory effort at the Ecosystem level.

Without the guidance of an all-encompassing vision that a sustained assessment or analysis group could provide, piecemeal progress will continue to be made by disparate Ecosystem elements, and these individual efforts will fail to have the necessary impact. There may continue to be duplication of effort, suboptimal allocation of limited resources, and inconsistent standards and metadata, all of which could have a stifling impact on the long-term utilization of data from NASA's missions. An overarching organizing body creating a clear set of well-defined and focused goals that can be communicated to NASA is critical to implementing the recommendations outlined in this report, and to ensuring that future planetary missions and researchers can continue to benefit from and stand on the shoulders of their predecessor missions and researchers.

As demonstrated by this report, the Ecosystem is an extensive network of disparate elements and almost certainly includes data producers, data providers, data preservers, and data users beyond those considered by this IRB. The scope of the Ecosystem is likely to expand over time as new instruments are invented, new types of data are considered, new analysis techniques are created, and new uses are imagined. A standing group that is chartered to consider the scope and needs of the entire Ecosystem on an ongoing basis and provide feedback and findings to NASA will ensure that NASA can respond to the needs of the Ecosystem and include those needs in its long-term planning. Existing AGs do not cover the breadth of the Ecosystem and are heavily focused on either very specific scientific goals or regions of the solar system or focus on very specific data types. The need for a more broadly scoped Planetary Data Ecosystem AG is clear.

A single IRB member provided the following **non-consensus** text: "In my opinion, neither the mandate nor the membership of the AG is clear. The mandate of the AG, as described, is not clear: aside from this being an organization with representation of a diverse community. The recommendation does not clarify which communities should be represented, except for stressing that selected communities like machine learning/AI, software development need to be included. It is not clear how disagreements between different communities will be resolved in making recommendations to the Planetary Science Subcommittee since it is likely that multiple stakeholders may have competing interests: and increased funding for one stakeholder can potentially cannibalize funding of another. There is a possibility the AG might actually slow implementation and create a deadlock in decision making if stakeholders do not agree on priorities. Also, there seems to be no discussion on whether the mandate can be achieved at the Program Scientist/Program Executive level at the PSD or by another NASA committee or Advisory Group. Most importantly, it is not clear whether such an AG will provide pointed and actionable recommendations for the PSD. Given the lack of clarity, in my opinion, an AG will not be useful at this juncture."

2.2.3. Increase Participation and Leadership in International Data Management Alliances

Finding 9 NASA is actively working to identify and share best practices, to leverage tools, to enhance the usability of data, and otherwise to reduce unnecessary duplication of efforts across NASA elements, both within the Science Mission Directorate and across the agency.

This IRB endorses this work and notes that adopting the Planetary Data Ecosystem concept would facilitate it.

Finding 10 The PDS is a leader in the International Planetary Data Alliance (IPDA). Both the PDS and IPDA benefit from this leadership.

Finding 11 Selective participation by Planetary Data Ecosystem elements in other, broader communities focused on data management and scientific information technology would be mutually beneficial.

Such organizations could include: the International Heliophysics Data Environment Alliance (IHDEA); the American Geophysical Union (AGU) Earth and Space Science Informatics section; expansions to or analogs of Earthcube; expansions to or analogs of the Federation of Earth Science Information Partners (ESIP), the International Virtual Observatory Alliance (IVOA), the Open Source Geospatial Foundation (OSGEO), the Research Data Alliance (RDA), and the World Data System (WDS). Organizations that bring together data providers, data consumers, tool producers, and funders, such as ESIP, are of particular value in enabling data preservation and use.

Several factors have limited planetary participation in these broader communities, including funding levels, the comparatively small size of the planetary community compared to others, and a lack of awareness in parts of the planetary science community of the value of data and tool interoperability across scientific domains, including Earth science, astrophysics, and others.

Recommendation 5 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.

The goals are to reduce the duplication of effort, to expand the interoperability of tools, including adoption and contributions to international interoperability standards, and generally to enhance the availability and usability of planetary data to advance NASA science priorities.

Recommendation 6 NASA should encourage collaboration around cybersecurity policies, practices, and infrastructure to preserve the integrity and availability of data and systems across the PDE.

Each ecosystem element has its own cybersecurity policies and practices, which often conflict with those of other elements due to differing levels of security. To ensure interoperability, the elements should be encouraged to discuss and align their cybersecurity policies.

Recommendation 7 NASA should maintain active leadership in the International Planetary Data Alliance (IPDA).

Recommendation 8 NASA should encourage the development of and participation in other cross-disciplinary organizations of data producers, data managers, and data users by PDE participants.

Finding 12 The PDS is internationally respected as a set of trusted data archives. However, none of the Ecosystem elements, including the PDS, has obtained any formal certification.

CoreTrustSeal certification is an internationally recognized formal recognition that is a key step in membership in the World Data System (WDS). Other data systems in NASA have found benefit in both the CoreTrustSeal certification and WDS participation. As the certification sponsored by the International Science Council, CoreTrustSeal is particularly appropriate for science data repositories. In addition, journals in many science domains now require that data be deposited with repositories that meet a certain level of rigor. The PDS nodes easily meet that standard; CoreTrustSeal certification would smooth the publication process. Repositories in other domains have reported that obtaining CoreTrustSeal certification is not onerous and that the process helped them to identify areas for improvement and to emphasize the benefits of following best practices.

Recommendation 9 NASA should seek CoreTrustSeal certification, and thereby WDS membership, for the PDS data nodes. NASA should encourage CoreTrustSeal certification for other PDE elements that serve as data repositories.

Given that PDS nodes are hosted by multiple organizations with management practices that differ at a level that is relevant to the CoreTrustSeal certification questions, the IRB recommends certification on a node-by-node basis for those nodes actively engaged in data preservation. Further, the NASA Earth Observing System Data and Information System (EOSDIS) achieved CoreTrustSeal on the equivalent of a node-by-node basis, although there are some historical reasons for that which may not be relevant to the present PDS. Should the PDS, after further examination and consultation with CoreTrustSeal, choose to pursue certification as a system of nodes rather than on a node-by-node basis, the IRB would consider that as meeting the intent of this recommendation.

2.2.4. Increase Interoperability Among Planetary Data Ecosystem Elements

Finding 13 Achieving the full potential for the Planetary Data Ecosystem will require increased use of common data and metadata standards across the Ecosystem and intentional efforts to reduce unnecessary proliferation of data and metadata formats, particularly those of limited use.

Finding 14 There is presently a significant quantity of valuable planetary data that do not meet the FAIR data standards.

Although the formalization of the Planetary Data Ecosystem is a relatively new concept, elements of this Ecosystem have been in operation for many years. These elements were created for different reasons, to address particular needs, and to serve differing, though often overlapping, communities. Achieving the full value of this Ecosystem will require investment in interoperability across these elements. The structural factors which inhibit software development, sustainability, and adoption noted in Section 6.1.2 strongly affect software for interoperability, as well as the development of standards and protocols.

There are also a multitude of factors that drive the proliferation of standards, a situation which is hardly unique to planetary sciences. There are clearly situations in which new standards are necessary, and it is rare that a single standard can serve the broad needs of something like a Planetary Data Ecosystem. However, for the Ecosystem to achieve its full potential, it will be necessary to cultivate a culture of interoperability and reusability and it will be necessary for NASA to create and use funding mechanisms that encourage interoperability, rather than unique or "one-off" solutions. It will also be important that the use of existing standards and protocols be addressed early in the mission design process, as missions have historically been a major source of standard proliferation in the planetary sciences, because mission teams have not prioritized the reusability of their data outside of the mission team.

Recommendation 10 NASA should prioritize the reuse of data and metadata standards, data format conversion tools, and Application Programming Interfaces (APIs) across other organizations rather than inventing new ones.

Taking advantage of experience that already exists in communities like the PDS, Earth science, astrophysics, and heliophysics as well as communities outside NASA and other government or research organizations could both improve interoperability and reduce the effort required in establishing data and metadata standards. By engaging these other communities, custodians of planetary data could open these planetary data sets to use with established software development and machine learning/artificial intelligence/advanced analytics (ML/AI/AA) methods.

Recommendation 11 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.

The FAIR self-assessment conducted by Earth Science Data and Information System (ESDIS) and presented by Ramapriyan and Behnke (2020) is an example of the relatively lightweight, but still informative, assessment envisioned here.

3 Planetary Data Stewardship

3.1. Introduction

The Space Studies Board Committee on Data Management and Computation was formed in 1978 and released a report in 1982 (NRC, 1982). Arising from recommendations in this report NASA formed the Planetary Data System (PDS). The Committee was "requested to examine the management of existing and future data acquired from spacecraft and associated computations in the areas of the space and earth sciences and to make recommendations for improvements from the point of view of the scientific user." This charge included several topics (the following is taken verbatim from the Committee report):

1. **Data System Planning:** The extent to which the complete data management and analysis associated with a mission is studied at the planning stage and its implementation is tied to hardware development.
2. **Preprocessing:** Consideration of the degree to which this may take place on board the spacecraft or at various locations on the ground.
3. **Distribution of Data:** The extent to which data should be examined by "quick-look" techniques (or even interactively) and the methods for improving the times required to supply principal investigators with their complete data sets.
4. **Data Standardization and Fidelity:** Recommendations for standardization of data formats, data compaction, error correction procedures, and other methods for assuring data quality.
5. **Software Development:** Recommendations for timeliness in software development, standardization of software procedures, and languages to maximize portability and inheritance in future projects and the improvement of programmer productivity.
6. **Distribution of Computational Capabilities:** The degree to which scientific users, depending on their computational requirements, can optimally utilize shared or dedicated computing facilities. This question will require an analysis of systems of different scale and should take into account projected trends in costs and capabilities of various systems.
7. **Mass Data Storage and Retrieval:** Recommendations for improvements in archiving, cataloging, and retrieval of data. This question should examine current practices and should make recommendations for standardization of data-base management systems, for the institutional siting of archives for the scientific user, and for desired technology developments to facilitate these three functions. Attention should be given to the problems of making catalogs useful and to convenient guides to data bases for both occasional and frequent users.

Consideration should be given to the expected trends in data storage and retrieval capabilities over the next one to two decades and their impact on future requirements.

8. **Interactive Processing:** An examination of the variety of ways in which human decision-making may be desirably introduced at various points in data-processing procedures.

It is remarkable how similar the charge from 40 years ago is to the charge of 2020's PDE IRB. It is also remarkable that several of the findings from the 1982 report (and again from the PDS Roadmap Study, 2017) match findings we are reporting here. Of particular note are several issues from the 1982 report, which will be repeated here: too little stakeholder involvement in data system planning; too little funding; responsible parties are often poorly identified if they are identified at all; documentation (of data, of the data catalogs, and of data-processing software and algorithms) is sometimes of poor quality; and users often have great difficulty determining which data are available and how to use those data; among several others.

This is not to say that nothing has been accomplished in the intervening 40 years. On the contrary, significant improvements have been made to all aspects of preservation of, distribution of, and usability of planetary data. The complexity of the problems being addressed have increased significantly, as have the expectations of data users. One obvious example of enormous strides made is this problem identified in 1982: "7(a) Often data must be purged from the archives in order to make room for current data." We are no longer purging data from the archives. The PDS, which grew out of the 1982 report, has done phenomenal work to ensure that NASA's, and the world's, planetary data are preserved. The reasons the same issues seem to arise time and again are complex and varied, but at least one of the reasons has to do with the difficulty in distinguishing what the PDS should be doing compared with what other elements of the Ecosystem should be doing, and there has not been a community group responsible for helping NASA keep data initiatives on track.

3.2. Findings and Recommendations: Planetary Data Stewardship

3.2.1. Planetary Data Stewardship Successes

Our first set of findings are related to the current state of the PDS and its role within the Ecosystem. These findings about the success of the PDS are only a few of many examples.

Finding 15 The PDS is a cornerstone of the Planetary Data Ecosystem and is critical to its success.

Finding 16 The PDS successfully preserves data, documentation, and expertise from NASA's planetary missions.

Finding 17 The PDS is a trusted, reliable source of planetary data, enabling data discovery, search, retrieval and analysis.

Finding 18 The PDS has been evolving to meet the archiving and preservation needs of planetary scientists and researchers, leading not just domestically, but internationally.

Finding 19 Within the limits of available funding, the PDS does a good job of meeting its primary mission of data preservation.

Finding 20 The PDS is working diligently to develop, manage, and maintain a complex set of data and metadata standards to meet a wide variety of planetary science data needs and expectations.

Finding 21 Efforts by the PDS software working group and individual PDS nodes to create PDS4 bundle creation and validation tools have been successful. Those tools facilitate the preparation and validation of data, and also improve the usability of data by the end user over the previous PDS3.

Recommendation 12 As NASA considers future evolution of the PDS, it should consider the positive aspects of what the PDS has accomplished within the context of the Planetary Data Ecosystem—as well as the context of history—and work to preserve the continuing positive outcomes from the PDS, including: maintaining or increasing funding levels as appropriate; working with Mission teams to continue to improve communications about and efficiency of data archival; continuing collaboration with domestic and international partners; and continuing to improve bundle creation and validation software.

3.2.2. Planetary Data Stewardship: Preservation vs. Processing vs. Distribution

Because NASA and its planetary sciences community (what we might call the Ecosystem) have relied so heavily on the PDS for so long, many aspects of the elements that do not fall under the PDS seem to have been left out or forgotten. This has led to a widely held idea that the PDS is the Ecosystem, and that the Ecosystem is the PDS. However, the PDS was not intended to be, and has never been funded to be, so comprehensive, and this new concept of the Planetary Data Ecosystem is needed.

Data archiving is a non-trivial activity that requires careful consideration. Creating discoverable data, processing “science-ready” usable data, and delivery of data to expert and non-expert users and programs and automated tools are all equally challenging and require different expertise and skill sets than those required to meet the goal of data preservation. The PDS Roadmap Study (2017) pointed out in its Finding II: “there is a mismatch between the services and functions PDS is equipped to provide and the very high expectations of its users and of NASA management.” In the years between

2017 and 2021 (this IRB), the mismatch between the PDS-funded services and community expectations have either remained the same or grown. As has been mentioned already (see e.g., Recommendation 2 and Recommendation 3), one path to resolving the mismatch between the funded services and community expectations is for NASA to explicitly define the scope of work for the PDS and other elements of the Planetary Data Ecosystem. Defining the full scope of work of the PDS and the rest of the Ecosystem was beyond the scope of this committee; additional community input is necessary, and it is important that NASA maintain the momentum generated by this IRB and institute a long-term broadly inclusive community assessment group to continue this work.

Finding 22 There is a mismatch between the explicit mission of the PDS (preserve data) and the perceived mission of the PDS by the broader user base (preserve, process, and distribute "science-ready" usable data).

Finding 23 The PDS receives direction from NASA that is sometimes outside the scope of the original agreements under which the PDS Nodes were funded

These "unfunded or underfunded mandates" have included the move from the PDS3 to PDS4 standard and archival requirements stipulated by the terms of PDART and other Research and Analysis grants.

Recommendation 13 NASA should ensure that PDS has adequate expertise and funding to maintain current standards and to support ongoing improvements, including funding of peer-review of data submissions.

Recommendation 14 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 prioritized 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.

3.2.3. Data Dictionaries, Taxonomy, and Peer-Review

With the development of the PDS4 standard, the Planetary Data System has become a *de facto* data standards organization. The PDS is working to develop, manage, and maintain these standards. Data providers to the PDS are required to meet these standards, including peer review of delivered data. Delivered data are reviewed by external reviewers, liens are generated, and data providers are tasked with dispositioning these liens. The general process is similar to the peer-review process of a journal publication.

At the present time, the PDS peer-review process is inconsistent in scope and veracity across the PDS discipline nodes, as there is no instructional documentation on the scope and content of the peer-review process available to guide the PDS nodes, the data providers, or the external reviewers in the specifics of the review. Additionally, there is no written policy or procedure for the peer review of node-created or -directed discipline-specific data dictionaries. It should be noted that mission specific data dictionaries may be peer-reviewed during the mission data peer reviews. However, there is no available policy or procedure for review of mission-specific data dictionaries.

Metadata is the fundamental method by which data becomes discoverable, accessible, and usable. During the archival process, data providers are required to use metadata ontologies to describe their data. In the context of the PDS, required metadata are described in a core data dictionary extended by discipline-specific and mission-specific data dictionaries. The PDS as a whole, with participation from representatives of the International Planetary Data Alliance (IPDA), develops the core PDS dictionary. The core dictionary contains the most common and widely used metadata classes and attributes. Discipline-specific dictionaries are developed by the individual PDS nodes or by interested community members.

The current set of PDS4 documentation (i.e. Data Providers Handbook) briefly describes the process of proposing or updating a discipline-specific dictionary. There is adequate documentation on the software process of creating a dictionary and submitting it to the PDS (see the PDS Dictionary). What is not obvious from the PDS documentation is the governance of metadata creation. There is no process specified for the overall architecture of the discipline dictionaries, how the dictionaries relate, who reviews the dictionaries for completeness and interoperability, or how the dictionaries relate to standards adopted by research community groups.

Finding 24 We reiterate Finding VIII from the PDS Roadmap Study (2017): “The PDS4 information model is well-documented at a highly technical level. However, there is a critical need for broader documentation and training for all levels of users.”

Finding 25 The PDS peer review of data and metadata for archiving is inconsistent across nodes and between reviewers. This includes peer review of metadata standards and documentation such as the discipline-specific dictionaries.

Finding 26 The PDS develops discipline-specific metadata dictionaries, and this is at the discretion of the PDS Nodes, with input from the data providers. It is not obvious that there is an overall architecture for these dictionaries, nor is there an established peer-review process to ensure interoperability or non-duplication of these dictionaries.

Finding 27 Data sets from the same mission that are preserved in separate archives with different data dictionaries and formatting standards are not always well-linked with one another to enable integral scientific studies.

Recommendation 15 All data dictionaries and information models for the PDS and for other archival elements need peer-review and contextual review (i.e., do these data dictionaries link well with other and existing data dictionaries while avoiding unnecessary redundancy?).

Recommendation 16 Create a shared, common taxonomy, controlled vocabulary, high level data dictionary, and/or glossary of terms across the Planetary Data Ecosystem. This will substantially advance the machine-actionability of Ecosystem data, and specifically improve interoperability and reusability as described in the FAIR data principles.

3.2.4. Planetary Data Stewardship: Governance

Because of the mismatch between the PDS' funded mission, the community's expectations for the PDS, and the roles the PDS nodes play within their respective communities, there can be miscommunication and misunderstandings both within the PDS and between the PDS and other elements of the Ecosystem.

Finding 28 The PDS is structured as a set of federated discipline-specific nodes. Although this structure has substantial benefits in meeting the needs of the differing scientific disciplines, it also sometimes leads to duplication of effort, unnecessary duplication of data, inconsistency in tools, fragmented access to data, and potential confusion for community members.

Some examples include Data Management Plans unique to nodes; node-specific archiving requirements; search mechanisms unique to each node; lack of systemwide cybersecurity standards and infrastructure; and ontology used to align various node metadata schema.

Recommendation 17 NASA should consider a more open and centralized Management Council for PDS governance that includes greater emphasis on systemwide governance in regard to structure, standards, and related processes. A major goal should be to increase the efficacy of decision-making and multi-way communication with Ecosystem stakeholders.

Recommendation 18 The makeup and distribution of nodes should be examined more closely to ensure that the PDS contains the appropriate and relevant node elements and subject matter expertise, that unnecessary duplication of effort and data do not occur, and that appropriate flexibility regarding scope and content is built into policy.

A more systems-engineering approach would include robust, standard metadata across the PDS with the capability to filter by node; standard search tools which provide cross-node functionality for all users. As the Planetary Data Ecosystem matures and additional elements (archives, repositories, needs, and users) are recognized and integrated, it will be essential to coordinate the metadata standards to promote interoperability.

3.2.5. Mission Teams Expectations

The concept of a Planetary Data Ecosystem is a new idea. The lack of an overall guiding hand to many of the Ecosystem's archival and other long-term needs has caused many inefficiencies, including: duplication of effort; untrained or informally trained mission archival managers sometimes re-creating others' mistakes; ill-informed planning of time and monetary costs for archiving; misunderstanding of roles; and other non-optimal approaches to mission and instrument data archival.

Finding 29 Mission teams are often required to re-format ancillary, NASA-produced data products for archiving.

There are several types of spacecraft data that are not directly produced by mission teams. These data and ancillary information are often produced by other NASA entities and are used by many mission and instrument teams. Since these data products are used by many missions, it is not practical, cost-effective, or efficient for data end-users for each mission to develop its own conversion to the PDS4 standard.

Recommendation 19 Mission teams should not be re-formatting NASA-produced data for archiving; this should be internal to NASA. It would make more sense for NASA radio-science experts to decide on a single, existing archival standard format for spacecraft tracking and ancillary data and to directly archive these data without relying on mission intervention.

Archiving of an instrument's raw and calibrated data is currently a required element in Planetary Science Division mission Announcements of Opportunity. During the proposal phase, mission teams define archival data products, data calibration methods, and data release schedule. Data archival requirements contained within Announcements of Opportunity are an excellent first step to the overall mission archival process.

Once a mission is funded, the development of a Science Data Management Plan (SMDP) is required, with a baseline due at the mission Operational Readiness Review (ORR) per NPR

7120.e and NPD 2200.1. The Science Data Management Plan is the agreement between the mission, NASA Program Management, and the Planetary Data System that identifies the scientific data to be archived, the archive in which the data will reside, and the schedule upon which the data will be released. Although missions generally meet this NPR 7120 requirement, the level of maturity of the design and implementation of archival and foundational data products varies considerably, and it is not uncommon for the need for a data acquisition change to be discovered during the mission in order to meet level one requirements. The variation in maturity can be attributed to the complexity of instrument build schedules and the length of time between launch and arrival at the target of interest. Issues in instrument build schedule can result in the delay of data product development. Long cruise phases offer a longer time window for data product development that may be appropriate.

Finding 30 Missions (and instrument teams) do not always plan data acquisition with archival or foundational data products in mind.

Long-term archiving of data for future infrastructure use (targeting of imagers and controlled products for landers, for example) requires standards for controlled, “foundational” data products and associated metadata.

Finding 31 The Science Data Management Plan is the roadmap to the mission archival process. However, there is no explicit link between this plan and the mission level one requirements in many missions. Without this direct link, there may be a contractual gap resulting in the non-delivery of the expected science data to the archive.

Finding 32 Changes in mission plans, including updates to data acquisition plans, changes in instrument build schedules, and delays in data product development, can often disrupt the original proposed Data Management Plans unless trained data management personnel are an integrated part of the science team.

Recommendation 20 NASA should review its contract agreements to ensure that mission instrument data archiving and future access and usability is an obligation that is appropriately considered and funded. As part of the agreement entered into with NASA, mission and instrument teams should be expected (and funded) to develop level one requirements that include raw, calibrated, higher-level, and foundational data product planning, execution, processing, delivery, and archival.

Recommendation 21 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.

Involving data archivists or skilled data management personnel early in mission definition and planning would ensure application of good practices for data archiving, but missions may not engage this expertise without NASA encouragement.

3.2.6. Data Archival Costs

Finding 33 Archival costs are difficult to estimate in the current state of the Ecosystem.

Typically, archival costs are distributed through many mission elements. Mission elements involved in the archive process include mission management, science team members, instrument team members, and science operations center members. These groups may not record time spent on archive tasks at a granular enough level for mission teams to have insight into the amount of time and effort the team engages in designing, developing, and implementing the archive. Additionally, it is unclear how the Planetary Data System accounts for reviewing, validating, and releasing mission archives or if this information is available to NASA management.

Finding 34 Archival cost estimates for Research and Analysis data producers are even less well constrained.

In many cases, would-be R&A data providers are new to the archival process. They do not have a reasonable basis for estimating the time and cost of data preparation and submission to the archive. The Planetary Data System spends significant time and effort to assist R&A data providers. It is not clear how the PDS accounts for this time, whether this time is currently funded, or if the information is available to NASA management.

Recommendation 22 NASA should consider a series of investigations or workshops to better understand the full costing of archival for various personas: mission archival managers, telemetry managers; instrument archival managers, R&A data producers; etc. The results of these workshops should be made publicly available and should be included with Data Management Plan templates.

Finding 35 Rigorous metadata requirements are an important requirement for data sets appropriate for inclusion in the PDS. These requirements also limit not only what can be added at a later point but also have challenged mission managers early in mission data planning.

Recognized within this finding is the increased burden and accompanying costs for PDS in trying to include data that do not fit well within the PDS element.

Finding 36 Some non-mission data providers who have little to no experience archiving with the PDS struggle with its mission-focused archival expectations, including costs related to archiving.

Finding 37 Training for archiving with the PDS has been conducted at training events offered by a few institutions, as individual training at academic and research institutions, or at the Data Users Workshop offered approximately every two years. However, this training is neither as widely available as it should be, nor widely known by members of the community or those wishing to enter the community.

Recommendation 23 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.

3.2.7. User Registration

Finding 38 A user registration system may be useful in better understanding the user community, increasing participation in the American Customer Satisfaction Index (ACSI) customer survey, and understanding use patterns across PDS nodes.

Tracking users by Internet Protocol (IP) address provides no contact mechanism and is becoming increasingly inaccurate given the increasing use of institutional egress proxies that put large numbers of users behind a small number of Network Address Translation (NAT) IP addresses, highly mobile users, and users working across multiple computers. The PDS also recognizes that some form of user registration may become essential, particularly in the context of public cloud computing, in order to manage the egress costs associated with any given user.

A lightweight user registration system could be of value to the PDS and the broader ecosystem. Such a system should be consistent with Identity Assurance Level 1 and Authenticator Assurance Level 1 as defined by NIST Special Publication 800-63 (NIST 2017). Such a system does not have to be comprehensive to still have substantial value. It should still permit unauthenticated browse of available data and images, such as the way EOSDIS presents Global Imagery Browse Services (GIBS) browse imagery. Such a system may, particularly initially, only affect the interactive download services, allowing large download services to function without authentication.

The 20-year experience with a user registration system within EOSDIS is relevant in evaluating the cost, benefit, and user impact of such a system, as well as best practices around user registration. Based on information provided to this IRB, the PDS has solicited input from other scientific data management systems that have implemented a user registration system to better understand the benefits and user impact. Additional outreach is planned, and the IRB endorses that effort.

Recommendation 24 **NON-CONSENSUS:** Several members of the review panel strongly recommend that the PDS move forward with a lightweight user registration system. Other members have concerns and strongly recommend a cautious approach so that any registration system implemented does not create additional barriers to access to, acquisition of, or usability of planetary data.

3.2.8. Cloud Computing and High-End Computing

Cloud computing is a term that can have multiple meanings to different people. Vendors of commercial services and the popular press have popularized a definition which refers to on-demand computing delivered over the Internet. However, cloud computing can also be understood to be more generic and include any form of computing infrastructure which is on-demand, with configuration and control done via software and automated methods. In this latter sense, cloud computing can be broken into public cloud computing (which is provided over the Internet, typically by a commercial provider), private cloud computing (where the underlying resources are dedicated to a single organization and are generally not accessible via the Internet), and hybrid cloud computing (a mixture of public and private cloud computing). Cloud computing can offer a wide range of advantages, particularly for organizations with highly dynamic workloads. In scientific computing, there is also a high degree of interest in public cloud computing as a means to allow users to bring the compute to the data, enable a wider array of data manipulation services, and generally reduce the need to download data to accomplish scientific goals, as noted in the PDS Roadmap Study (2017).

This review board was provided with a copy of the PDS cloud computing strategy, which is not a publicly accessible document. The review panel also interviewed some of the people who wrote this strategy. The board suggests that this strategy should be made public and kept up to date.

The PDS cloud computing strategy recognizes the value and roles of different computing modes, such as public cloud computing, hybrid cloud, private cloud/high end computing, and user-provided computing. It recognizes that cloud computing, particularly public cloud computing, is a tool to be used appropriately, rather than an end in and of itself. It properly acknowledges the risks which public cloud computing brings, particularly to organizations which are subject to the Anti-Deficiency Act (ADA 1982). The strategy includes work to reduce technical debt and implement technologies, such as containerization, that will facilitate migration to cloud services.

There are advantages to using tools which can rapidly scale to meet user demand, and enable the rapid deployment of new services, but there are also cost risks associated with that demand and those services, particularly for organizations like NASA where additional user demand is not associated with business revenue. Cloud computing is an approach that enables analysis in place for planetary data and reduces the need for users to download data in order to manipulate it. As such, cloud computing may have value for planetary science. However, as noted above, cloud computing is an implementation strategy to be used in addressing user needs or achieving project technical objectives, rather than an end, in and of itself.

Finding 39 The PDS cloud computing strategy is reasonable, within the limits of available funding. The review panel suggests that PDS make this strategy public, in the interests of increased transparency with PDS users.

Finding 40 Putting data into a public cloud infrastructure is a valuable approach for enabling users to bring algorithms to the data and advancing a wide range of advanced computing methods, including machine learning and advanced analytics methods. However, data usability and data interoperability issues must be addressed first in order to realize the benefits of these advanced computing methods, as discussed in Section 5.

Finding 41 Like public cloud computing, high-end computing (high-performance computing) is an enabling technology for a wide range of advanced analytics methods, including machine learning and artificial intelligence. As with the previous finding, we find that data usability and data interoperability issues must be addressed before a subsequent review board or assessment body can recommend any changes to the high-end computing strategy as applied to planetary science.

Recommendation 25 NASA should continue to execute against the PDS cloud computing strategy, including selective refactoring of current systems to enable cloud migration, such as the adoption of containerization and further work to establish well-defined and well-documented application programming interfaces.

Recommendation 26 The PDS should continue discussions and collaborations with other NASA elements, including EOSDIS and OCIO, to leverage the work done in these organizations and ensure that Planetary Science Division needs are appropriately considered in establishing NASA standards and practices.

Finding 42 Public cloud computing can enable cross-organizational collaboration and should be considered as part of the strategy for further fostering the Planetary Data Ecosystem.

Finding 43 Based on the experience of other organizations, particularly EOSDIS, public cloud computing can have benefits for the Ecosystem beyond dynamic computing and the other items noted in the PDS cloud computing strategies. Public cloud infrastructure can provide for a common environment, enabling collaboration across the PDS nodes and even across other Ecosystem elements through what amounts to a common credential infrastructure. This can help increase communication across Ecosystem elements, in addition to enabling common tools and reducing unintended duplication of efforts.

Finding 44 Although the various nodes of the PDS are important for serving the different scientific focus areas of the planetary community, the organizational segmentation has resulted in some duplication of effort, some tools that are similar but not necessarily interoperable, and systems more tied to organizational structure than may be optimal. A common cloud infrastructure could enable common infrastructure while allowing nodes to maintain their scientific domain foci.

Recommendation 27 NASA should consider the impact of cloud computing adoption on organizational efficiency and the development of a broader planetary ecosystem, above and beyond the technical capabilities that public cloud computing brings to addressing data provider and data user needs.

4 Systemic Barriers to Data Preservation

4.1. Introduction

The following section presents a collection of findings and recommendations that address unmet preservation needs within the Planetary Data Ecosystem. Specific discipline-dependent findings and recommendations result from examination of responses to the Request for Information (see Appendix E), discussions within IRB subcommittees, presentations by invited expert speakers, and board members' individual experiences. The composition of data sources available to the IRB directly influences the findings presented. As such, these findings are representative of the input sources, and are not a complete set of possible data types that are not represented in the PDE.

4.2. Findings and Recommendations: Barriers to Data Preservation

4.2.1. Preservation of Data Not Within the Purview of the PDS

Finding 45 Some types of mission data do not have a permanent, designated home.

Finding 46 The Planetary Data Ecosystem has unmet data preservation needs in areas outside of mission data.

Finding 47 There is a lack of systematic guidance on which data are appropriate for which archive/repository, and the level of curation effort required for that class of data. Not all Ecosystem data that should be preserved are appropriate for the PDS.

As discussed in section 3.2.1, the PDS is a cornerstone of the Planetary Data Ecosystem, focusing primarily on mission science data. The PDS does an admirable job of tending these data. The IRB subcommittees, through RFI submissions, expert reports, and individual experience, have identified general classes of planetary data that do not have specific, permanent, usable archives or repositories. The following classes of data are not exhaustive, but illustrative of entire classes of data that are collected or hosted on an ad-hoc basis, rather than in a systematic way:

- Telescopic observational data
- Returned sample analytical data
- Analog field site data
- Mission operations and contextual data
- Historical information
- Higher-level mission or R&A-produced data products
- Software

- Simulation data

These information types greatly benefit the community, especially interdisciplinary studies, but are not adequately preserved because NASA has not identified specific archives/repositories or standards for, nor have they levied systematic preservation requirements on, these broad classes of data. Some of these data may belong with data archived with the PDS, but others may belong in a different repository. Although this may seem like mostly an archiving issue, it also reflects the conundrum that users do not know what kind of data they can expect to find within the Ecosystem. Users also do not know if searches are unsuccessful because the data do not exist in the public sphere, or if they lack the ability to find it.

For a myriad of reasons, some data providers do not meet the PDS standards of science and data product peer review. They turn to other repositories (e.g. Dataverse, academic publishers' websites including supplemental material, GitHub, Flickr, etc.). These repositories have lower levels of trust, discoverability, interoperability, and permanence. The differences in repositories' level of compliance with FAIR data principles (see Section 2.1.4) may be acceptable depending on the specifics of the class of data to be preserved.

The following recommendations are applicable to the broad topic of unmet preservation needs. Subsequent findings and recommendations in Section 4 cover a few specific classes of data to illustrate the many unmet data preservation needs. Together, these examples suggest the potential value of a Planetary Science Division-wide strategy for archiving and data preservation, with emphasis on data discoverability, accessibility, and usability.

Recommendation 28 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.

Recommendation 29 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.

Recommendation 30 The Science Mission Directorate should elevate support for information and data science issues to parity with other areas in order to systematically address NASA's unmet data preservation needs.

A way to achieve this is by including information and data science personnel assignments at higher administrative levels. These assignments will also ensure that barriers to access and usability are addressed (see Section 5).

4.2.2. Preservation of Telescopic Observational Data

Finding 48 Some telescopic planetary observational data are not preserved accessibly.

Arecibo Observatory data are not, generally, currently archived in a PDE element, though some data are archived in the Small Bodies Node (SBN) of the PDS. This is also true of other ground-based radar telescopes. Ground-based planetary radar observations will expand in the next several decades in support of NASA and NSF mandates, with the Green Bank Telescope anticipated to be operational in a few years, a potential new vision for Arecibo, and upgrades at Goldstone to respond to specific needs in planetary science and defense. Arecibo data processing tools and software are not archived, and higher-level processing is science-need-dependent. It is not realistically possible for a scientist who is new to accessing Arecibo data sets to process them to the appropriate levels.

Some major ground-based observational facilities do not currently maintain public archives, even though it would be very useful for the planetary science community to be able to access those data in order to, for example, search for serendipitous detections of small bodies. Examples of such facilities include Palomar Observatory, Las Cumbres Observatory, Lowell Discovery Telescope, TRAPPIST, and others. In principle, however, any observational facility that obtains imaging data has the potential for serendipitously observing small bodies, and so should also be considered within the scope of this finding. The preservation of pointing-coordinate logs by observational facilities along with field-of-view and orientation information are critical to facilitate the discoverability of serendipitous detections of small bodies. Because knowledge of the orbits of small bodies improves with time, searches of archived data for small bodies not identified at the time of observation are often fruitful and provide significant improvements to their orbits and properties. In order to enable these searches, it is critical to know when and where a telescope was pointed, and what the shape and orientation of the field of view are. Additionally, not all telescopes that maintain public archives provide analysis-ready data, inhibiting use of that data by users who may lack the expertise and/or software to perform the needed data reduction.

Recommendation 31 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.

Recommendation 32 NASA should support the establishment of public archives of analysis-ready data from observational facilities for which such archives do not already exist.

For many of the facilities relevant to Recommendation 34 that are operated by private and/or foreign entities, NASA will not have the authority to unilaterally establish public archives for those facilities. Instead, examples of alternative ways that NASA could support or incentivize the creation of public archives of analysis-ready data for such facilities include agreements for NASA to process and archive all data by itself (e.g., as part of PDS operations), the direct funding by NASA of archiving efforts by observing facilities themselves via contract awards or cooperative agreements, or shared-cost partnerships between NASA and the observing facilities to create the desired archives. Given that it will not be possible to create public archives of analysis-ready data for all relevant facilities due to finite resources, NASA (through the PDE-AG) will also need to identify the highest priority facilities for which public archives should be created, ideally based on input from the planetary science community and cost-benefit assessments of potential scientific impact of each archive.

4.2.3. Preservation of Returned Sample and Analog Data

Finding 49 Returned sample contextual and analytical data are not preserved in an organized fashion.

Physical samples from non-terrestrial sources are becoming more numerous, with forthcoming increases in the number of acquisition attempts and in total mass samples curated. These samples are curated by Johnson Space Center’s Astromaterials and Exploration Science (ARES) Division. NASA maintains a rigorous set of requirements for the curation and management of these materials. However, there is no requirement levied upon sample return missions for the archival of mission-supported laboratory analytical data. Additionally, there is no agreed-upon standard for metadata describing geochemical or laboratory analytical data derived from the returned samples, although some work toward a standard is in progress. In the case of sample return missions, the data generated by preliminary mission-supported analysis of returned samples should be treated similarly to that of data returned by the spacecraft. Subsequent laboratory analysis data produced by later analysis should be preserved, but possibly in a secondary repository with reduced barriers to entry.

Recommendation 33 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.

Recommendation 34 NASA should require data preservation with appropriate metadata in an approved archive or respository for data produced by laboratory analysis of returned samples supported by ROSES Data Analysis Programs (DAP).

Sample collection contextual data do not currently have a clear link to other data in the Ecosystem. The first type of contextual data to be considered are the environmental conditions during the collection of samples on planetary bodies. There is clearly an impact of planetary environmental conditions, e.g. temperature, pressure, humidity, etc. on scientific observations and samples. However, it is not clear whether the environmental variables that influence the scientific observations are preserved in metadata that is archived alongside the scientific data to convey the proper context in which the scientific data were taken. Care must be taken to ensure that linkages exist in the preserved data to tie the contextual data to the returned samples.

Recommendation 35 NASA should adopt or develop a standard set of metadata and links to ensure that contextual data are adequately tied to returned and gathered samples. With Mars 2020 gathering and caching samples for later return to Earth, time is of the essence.

Finding 50 Planetary analog data sets do not have a primary repository.

Analog field data do not easily fit into just planetary science or just earth science data archives. These data are, in many cases, hosted by individual researchers or in disparate locations with no guidelines or standards for archiving. The lack of standards hinders accessibility and discoverability of these data. Analog field data may include physical and digital products including personal field notes, photographs, audio, video, and post-field laboratory measurements performed on samples. Establishing a single-host service that provides archiving guidelines to serve analog data and/or to collect and harvest links to data from other hosts of planetary and earth science data would enhance accessibility and discoverability of past and future data products.

Recommendation 36 NASA should assess the current state of planetary analog repositories and develop the requirements for the establishment of a permanent planetary analog repository or archive.

4.2.4. Preservation of Mission Operations Data

Finding 51 Mission operations data are not preserved in an organized fashion.

Engineering telemetry from planetary missions (both from the spacecraft as well as payload instruments) is a critical and irreproducible data set that currently has no systematized archive or repository. Telemetry is received by the Deep Space Network (DSN) and routed to mission team spacecraft and science processing facilities. Telemetry is stored for a finite time at the DSN, but the full telemetry stream is not systematically preserved or available. Future planetary missions and those currently in development could benefit tremendously from having access to engineering telemetry from relevant previous missions. For example, engineering telemetry from previous missions like Cassini and Galileo could be of great value to current mission designers, planners and engineers on the Europa Clipper project to guide development of the operations plan and perform technical trade studies. We do recognize that there may be some limitations to access based on regulatory considerations.

Recommendation 37 NASA should establish a primary archive or repository for mission telemetry streams that is accessible to the planetary science community to the extent permitted by regulatory limitations.

Mission operations and planning information are also critical and irreproducible data sets generated by mission teams. As with mission engineering telemetry, there is not a systematized archive or repository for this information. There is no explicit requirement in current mission Announcements of Opportunity to archive this information. The community has made attempts to capture this information on an ad-hoc basis. However, there is a lack of consensus on how to include this information in PDS4 mission bundles, or if PDS is the appropriate place to archive this information. Currently, much of this class of information is stored at mission operations centers (e.g. JPL, APL) and is not readily available to non-company personnel.

Recommendation 38 NASA should establish a requirement for the preservation of mission operations and planning information that makes the information accessible to the planetary science community to the extent permitted by regulatory limitations.

Finding 52 Historical information is not preserved in an accessible manner.

There is currently no place to archive historical data, press releases, internal communications, etc., that might be useful for archeological studies of NASA missions past (and present-that-will-become-past). The bulk of the data that are currently found in the PDS archives consists predominantly of scientific measurements from spacecraft or ground-based instruments, as well as the associated metadata for these measurements. As these are the primary outputs of planetary missions (which represent a significant fraction of the NASA Planetary Science budget), and data archiving is required for these missions, it is not surprising that they dominate the archives. However, there are other sources of data that are important for planetary science investigations that are not represented in the present archives. Press releases, communications, technical memos, or other items may be of interest to future archeological studies of NASA missions or projects. Currently, NASA Headquarters and NASA Center archives and libraries do preserve some historical records, press releases, internal communications, etc., but in a manner inaccessible to the public. We recognize that each of these types of data have their own unique challenges associated with long-term preservation and ensuring that archived versions comply with FAIR data standards.

Recommendation 39 NASA should evaluate and develop a plan for historical information preservation with the aim of making these data available to the public to the extent possible.

4.2.5. Preservation of Higher-level Data

Finding 53 Higher-level data products are not systematically hosted or supported by NASA.

The PDS has often been discussed as a location for higher-level data products, possibly due to its preeminence in planetary science as the primary archival location. However, PDS has not been provided with adequate support to ingest or host these data products. As discussed in the PDS Roadmap Study (2017) Finding XIII: "Higher-level data products [...] are not always included due to lack of resources needed by missions to complete the archiving process." Instead, high-level data products derived from mission data or the synthesis of mission and analog data are currently supported in a myriad of locations outside the PDS nodes. NASA has not clearly designated a preferred location(s) for other higher-level data products, including analysis-ready data, that would support machine learning/artificial intelligence/advanced analytics methods and other uses.

Recommendation 40 NASA should establish requirements that specify the archive(s) or repositories of record for higher-level data products, with the ultimate goals of systematic collection and reuse of these high-level data products.

High-level data products created by research and analysis (R&A) scientists and other researchers and data providers, including amateurs, are often generated after a mission has ended. These products are not necessarily appropriately archived under the PDS. Individual R&A data sets that are too costly or too troublesome for their producers to archive with the PDS will often end up stored on thumb drives, DVDs, CD-ROMs, tape drives, hard drives, floppy discs, optical drives, etc., on a shelf in the data producer's office, rather than being archived. Although the goal of archiving R&A data with the PDS may be a lofty one, NASA should understand that the pursuit of ideal solutions may sometimes leave the Ecosystem with no solution rather than a solution that is sufficient.

Recommendation 41 NASA should establish guidelines for preserving high-level data sets of interest that are not appropriate to PDS archiving. Designate data repositories that comply with FAIR (Findable, Accessible, Interoperable, Reusable) data principles.

The ESA's Planetary Science Archive's Guest Facility may be one model for how this can be accomplished.

4.2.6. Preservation of Software and Models

Finding 54 Current policies governing the preservation of software and planetary data processing pipelines are inadequate for ensuring their discoverability, accessibility, and usability.

The ROSES 2021 PDART Appendix uses the term “archive” to describe the action of placing code on GitHub. This is an inappropriate use of the term “archive” in this context. GitHub is not an archive (see Section 2.1.3) because it does not ensure long-term preservation and has lower bars for required documentation than true archives. Instead, use of the terms “deposit” or “working repository” would be more appropriate. Although seemingly minor, this use of the term “archive” to refer to GitHub in an official NASA document undermines efforts to create genuine archives of high-value NASA tools.

Meanwhile, the PDART program requires principal investigators of selected proposals that center on the development of software to deposit their software in a NASA-managed GitHub repository, but this requirement is not enforced. Some, but not all, NASA-funded software may be found on various top-level NASA GitHub repositories, or code.nasa.gov, while other code may only be made available via individual authors’ GitHub repositories, personal websites, etc. This wide variety of “habitats” can also complicate the ability to find tools that would suit a particular purpose, as they may not be amenable to being found via search engines as top results using common, non-targeted search terms. Furthermore, since software development is often considered to be of secondary importance to data acquisition, when individuals do build tools, they are not typically as well advertised or announced to the community as scientific research results are. Scientists cannot use tools that they cannot find. See also Section 6, which discusses structural factors that limit the development, sustainability, and adoption of software.

Recommendation 42 NASA should develop a comprehensive software preservation and archiving strategy that ensures discoverable, accessible, and usable software tools. The curation of the collection of NASA-funded software products through a designed software node within PDS, a centrally managed catalog, or with another approach will ensure the successful implementation of NASA open-source software policies.

Much of the data that are archived in the Planetary Data Ecosystem represents the end state of a long series of data processing steps. This is particularly true of data from spacecraft missions, where instrument data are compressed, passed through a downlink system, unpacked, and merged with metadata before even reaching a state that would be considered “raw.” These raw data then undergo processing by the operations and science teams to create the desired higher-level data products used for scientific analysis, often with many intermediate levels of processing recorded. Inherent to these processing pipelines are choices made by the teams for analysis and calibration steps that are often opaque to external researchers. These choices, which may have been the best possible at the time, could potentially be improved upon as our knowledge of the data, instrument, and object being observed improves. It is also possible that the pipeline code contains software bugs that have a minor but non-negligible effect on the output and thus resulting analysis.

Although teams often publish idealized algorithms or processing flow charts, significant differences can exist between what was intended and what could be or was implemented. Therefore, it is important for reproducibility that the software used to process data be archived in the configuration and state they were used in. While some individuals and teams make use of software repositories (e.g., GitHub) for software distribution, these repositories are not suitable as archives. These repositories have no guarantee of long-term preservation, comparable to that of the archived data. Additionally, repositories have minimal requirements for the documentation that would be necessary to understand the software, especially in the future when preferred software languages and packages have changed.

In general, given the diverse “habitats” (as discussed above) in which publicly available code can be found, requirements that NASA-funded software developers associate persistent identifiers (such as DOIs issued by organizations like Zenodo) with their published code would facilitate the management of a centrally managed catalog of NASA-funded software, help authors and users cite software, and allow systems that enable tracking of documents to also track NASA-funded software.

Recommendation 43 NASA should develop requirements for the maintenance of mission data processing pipelines so that non-team members can produce data identical to the output from instrument team processing pipelines.

Finding 55 Theoretical models, as well as model inputs, parameters and output data are not currently preserved.

Data generated by numerical models or other simulations of planetary bodies and processes are not currently preserved in a robust way. Examples of this type of data include N-body simulations of small body orbital evolution, smoothed-particle hydrodynamics (SPH) modeling of impacts into planetary surfaces, simulations of the atmospheric conditions of a specific place and time on Mars, or simulations of the evolution of lunar magnetism during its formation. Challenges associated with these data sets include the very large data volumes, the inconsistent availability of metadata necessary for their interpretation, and their dependence on the specific software version or instances. The PDS has been working on addressing some of these challenges including how permanent model results should be. Certainly, when a model is replaced by a newer version, a detailed assessment of the older model and input and example output should be archived. The specific model and modeling team may have different preservation needs.

Recommendation 44 NASA should develop a plan for the preservation of models and modeling data beginning with requirements for how models and modeling data should be preserved and linked to other Ecosystem elements.

5 Barriers to Access and Usability

5.1. Introduction

We define a successful planetary data ecosystem as one that efficiently and effectively supports FAIR data preservation, data access, and data use. Of course, best practices in supporting all three aspects of a successful data ecosystem vary depending on the data product in question and planned reuse. To date, the Planetary Data System (PDS), a key component of the broad concept of the Planetary Data Ecosystem (PDE or the "Ecosystem"), has primarily focused on long-term archiving and providing access to planetary data (see historical discussions in the PDS Roadmap Study (2017). Other components of the Ecosystem are less standardized. Some have limited (non-public) access.

Data users' needs and expectations have recently evolved to include expectations of higher-level data products and access to various parts of the PDE. As discussed in the PDS Roadmap Study (2017), users are additionally seeking services that "help them explore, discover, sort, and visualize data before they even download the data locally." Supporting increasingly variable data use is a major challenge for the PDS for a variety of reasons, including programmatic directives and funding provisions.

Data use and reuse are built on the prior tasks of archiving and accessibility. Data use frameworks envision providing FAIR (findable, accessible, interoperable, reusable) data access (see section 2.1.4). The FAIR data framework ultimately prioritizes the reuse of data. While current access and usability needs of users vary widely, this can be considered in comparison to the history and current goals of the PDS' primary focus on providing "a long-term archive of digital data products" (pds.nasa.gov). The PDS is the primary location for all NASA-supported planetary data, and it is currently undergoing significant updates in regard to usability. However, it is not currently provided support to prioritize data use, which is an essential part of a healthy and sustainable data ecosystem. The following section focuses on the potential for enhanced data use and reuse in the broad Planetary Data Ecosystem in addition to the PDS. Below, we have attempted a preliminary assessment of the data user community and their needs.

5.2. Findings and Recommendations: Barriers to Access and Usability

Finding 56 The high number of points of entry among and within Planetary Data Ecosystem elements does not meet user needs for searching and locating data.

For any given science use case, it is not always clear where to begin a search for science data. Many users resort to a generic search engine as a first step, which is not ideal. Even for data within the PDS, about 48% of PDS users survey respondents also indicated that a Google search was a primary tool in addition to manually searching PDS archives (PDS, 2020). An activity conducted by the Searching subcommittee exploring PDE user personas (see Appendix C) demonstrated that people searching the broader Planetary Data Ecosystem are even more likely to begin with a generic search provided like Google.

Unfortunately, NASA websites that host the official or most current versions of data sets often do not show up at the top of searches via Google or other search engines. Landing pages for Ecosystem data sets are typically not designed with search engine optimization in mind, so Google searches for data can be laborious and often fail to yield the desired data at all. The top result is most commonly Wikipedia. It would be useful to have data, documents, software, tools, and other supporting NASA science data information accessible from a small number of portals that could assist in the location and retrieval of NASA science data. We note that the PDS Roadmap Study (2017) articulated a similar finding for PDS: “There is a need for PDS to both expand and deepen its search services, with a view to making it easier for users to find and execute the search appropriate to their query” (Data Discovery, Finding III).

Recommendation 45 NASA should provide and advertise a better point of entry (or several well-connected portals) to its data, suitable for the broadest range of users looking for planetary data.

The goal would be to provide a small number of common and well-known pathway for users to start their navigation and to search for NASA science data, assisting both new users and more sophisticated users to find and use tools appropriate to their science investigation. Features of such portals might include:

- User Experience-informed design, help screens, and a chat bot for FAQ answers;
- Introductory material and contact information for all elements;
- Links to the various elements, so users can drill down to specific data collections; and
- The involvement of marketing/communication expertise in improving the search engine optimization (ranking in search results) of the most useful Ecosystem entry points, with or without direct collaboration with search engine companies.
- Leveraging work from other science disciplines for optimizing the results of generalized Internet searching, such as the work described at science-on-schema.org and including schema.org metadata in data set landing pages. This also has the benefit of enabling more efficient automated harvesting and interpretation of metadata to support a wide range of machine-actionable data access pathways.

Finding 57 Many Planetary Data Ecosystem data sets are difficult to browse or preview.

Users expect highly interactive search experiences as opposed to a static storage archive. Instead, users often must download and manipulate data sets before they can find out whether they contain desirable information. Effective, user-friendly search tools are limited or unavailable for most data sets.

Recommendation 46 The user search experience needs to be improved across the Planetary Data Ecosystem. PDE elements should partner with a user experience (UX) expert to understand the principles and guidelines for UX.

Finding 58 The endpoint of a search is often not the best version of a data set for the user's needs.

Higher-level data sets that are archived are usually not discoverable from lower-level versions of the same data sets. By and large, derived data published outside the PDS do not link to lower-level versions which may be archived outside the PDS, nor to data sets within the PDS. The trail can frequently get lost along the search process. The PDS4 standard includes powerful mechanisms for providing pointers between data sets and products, but application of these mechanisms has only started recently, is very incomplete, and is irrelevant for data sets archived outside of the PDS standard.

Recommendation 47 NASA should support and encourage expanded use of DOI-like identifiers for data, thereby connecting data at various levels of processing to assist users in locating the best version of a data set for their needs.

Use of such identifiers by both data producers and data users would establish clearer links between raw or analysis-ready data and downstream data/science products (e.g., Minor Planet Center (MPC) astrometry, journal articles), facilitating more detailed follow-up analyses and reproducibility studies. Designated entities responsible for assigning such identifiers could also be tasked with exploring and pursuing options for maximizing the ease and efficiency of data identifier use. One option would be to assign identifiers to “collections” of data (subject to the constraints of PDS3 and PDS4 data formats). This would avoid problems potentially caused by assigning identifiers at the data bundle level, which can be unwieldy, or by assigning identifiers at the full data set level, which is not helpful in cases where the data being referenced comprises only a small subset of the full data set.

Finding 59 Many other types of data are relevant to planetary science but are governed and archived by organizations in disciplines other than planetary science.

The rapid expansion of exoplanet data acquisition over the last decade has resulted in remarkable synergies with planetary sciences. In addition, data from astrophysics, heliophysics, earth sciences, life sciences, health sciences, the Human Research Program (HRP), and sample collection on other planetary bodies are becoming more and more relevant to each other. A few examples include:

- Solar wind measurements from heliophysics missions which are input conditions for interpreting planetary atmospheres.
- Earth science observations using identical techniques to remote sensing of other planets providing ground-truth comparisons.
- Astrophysical surveys that provide serendipitous observations and measure properties for asteroids and comets in their fields of view.
- Laboratory analyses of extraterrestrial materials using the same equipment and techniques as terrestrial samples.

Cross-disciplinary interrelationships are not new, and many have been the subject of significant study in the literature. However, the volume of data that exists at these boundaries between disciplines is changing. Exoplanet characterization is accelerating, and necessarily will look to planetary science for interpretation and analogs, while planetary science investigations will use upcoming exoplanet research to give context to the formation conditions of our solar system. Telescopic surveys on the ground and in space, such as the Vera Rubin Observatory and the Nancy Grace Roman Observatory, are set to increase data volumes by many orders of magnitude over what is currently analyzed by the community. Sample return missions are ongoing or planned for multiple different planetary bodies that will require new analysis techniques and Earth analog comparisons.

The data that results from evaluating the effects of radiation on the health of astronauts in deep space missions, as in the Human Research Program, is archived in the Life Sciences Data Archive (LSDA) at Johnson Space Center. In the same missions, images are taken, rock samples are collected, and the effect of radiation on the operation of spacecraft microelectronics is evaluated, resulting in various data sets that are preserved in archives other than the LSDA, mostly PDS. If an integral study of data collection in a particular mission is to be conducted, the various data sets that have been preserved in different archives all become important for review and analysis. Some interoperability or linking between the different data sets becomes necessary for easy searchability of the mission data.

In many of these cases, there is a small up-front cost to ensure that the necessary metadata are recorded and published to enable these new data sets to be usable for planetary science studies and comparisons. Attempting to construct these tools after the missions and data structures are defined usually has a much higher cost and can be more disruptive to the ongoing mission.

Recommendation 48 NASA should continue to support non-planetary data archives and encourage cross-communication between planetary and non-planetary metadata developers.

It is important for the Planetary Science Division to actively engage data producers in other divisions early in mission development to maximize the utility of these data sets for planetary applications and minimize development costs.

Finding 60 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.

Planetary data sets are rarely accompanied by metadata describing their quality. This lack of metadata makes it difficult for users to assess whether a data set is suitable for a particular application. Many data sets are archived in raw format with documentation about radiometric and geometric calibration methods. The usability of this documentation is highly variable. Most users lack the time, skills, and the software needed to calibrate data from raw data sets. Older data sets are not only difficult to find but are effectively inaccessible to most users. The NASA PDART program is currently addressing these issues, but they are likely to persist, hampering usability and limiting longitudinal studies.

Recommendation 49 NASA should fund the development of more analysis-ready data (ARD) products derived from the lower-level products created by NASA missions.

The Committee on Earth Observing Satellites (CEOS) has developed a flexible, extensible definition of analysis-ready data, and is actively developing specifications for ARD products from optical and synthetic aperture radar observations of the Earth. NASA should consider using the CEOS ARD framework as a starting point and extending it to apply to planetary data sets. However, we acknowledge that the starting point of the CEOS ARD framework is very far from the finish line for a planetary ARD framework. Earth Science data products are all relative to a single, well-surveyed target body. Addressing the differences in target type, coordinate systems, atmospheres, surface conditions, and other factors for the wide range of planetary data will require an enormous amount of work.

Finding 61 Many Planetary Data Ecosystem elements are designed for an expert group of users but should serve a broader user base.

As described in Section 2.1.2, the user base for the PDE elements is broad. Stakeholder communities tend to be confined to individual Ecosystem elements, yet the elements frequently share user groups. Community outreach, education, and training components vary widely across elements. There is little community outreach that fosters open science or “invites” new users via easy access to, and discovery of, new data.

The PDS and other archives tend to have steep learning curves. Language on websites employ terminology and jargon that can present a significant barrier to new users. Many data archives and repositories do not have the requisite tools to facilitate the search and location of data for even the more experienced users. They are often (though not always) designed by small, connected, expert communities for use by those communities. Many Planetary Data Ecosystem community members, or those who would like to join the community but are prevented by knowledge barriers, have significant challenges in discovering and accessing these data sets. We note a similar finding from the PDS Roadmap Study (2017) Finding VIII: “The PDS4 information model is well-documented at a highly technical level. However, there is a critical need for broader documentation and training for all levels of users.”

Recommendation 50 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.

Possible strategies include:

- Hosting workshops to improve insight into Ecosystem user community needs and to capture a working list of needs related to data discovery and access.
- Developing user-friendly guides, video tutorials, and online classes (either live or asynchronous).
- Involving science communication experts in writing or editing documentation and website language geared towards newer users.
- Taking advantage of materials that have already been developed by public engagement, media information, and education offices to educate the media and public about Ecosystem elements. These materials provide both background and current findings related to mission data and resulting scientific analysis and findings, written in plain language.

6 Barriers to Development

6.1. Findings and Recommendations: Barriers to Development

6.1.1. Multiple factors inhibit the development and use of machine learning, artificial intelligence, and advanced analytics (ML/AI/AA) methods in planetary science.

This IRB was specifically charged with evaluating the state of planetary science data in the domain of ML/AI/AA methods. We also include advanced visualization approaches in this domain, particularly those visualization methods optimized for image analysis and image enhancement.

ML/AI/AA methods are often associated with so-called “big data” (that is, data sets that are too large to analyze by traditional methods). They are well suited to working with large quantities of data because many of these methods (particularly machine learning) work best when performed on large data sets. Planetary science, unlike other fields, has only recently begun to be faced with data sizes that are simply too large to manually analyze.

Finding 62 ML/AI/AA methods have critical roles in mission planning and execution, as well as in maximizing scientific value from mission data.

The review board identified many successful and valuable applications of ML/AI/AA in planetary science, as identified in the responses to the Request for Information (Appendix E), interviews with experts, the experiences of panel members, and our survey of the literature.

Given the latency and bandwidth limitations of the Deep Space Network (DSN), planetary exploration presents extreme cases for what is presently termed “edge computing,” where analysis and decisions need to be done by the device where the data are collected, without the data being sent to a central processing system. An example of edge computing is the autonomously guided entry, descent, and landing of the Perseverance rover on Mars. There are also many other applications, including the use of edge computing to prioritize data for downlink through the DSN. ML/AI/AA methods are widely recognized as being particularly applicable to these types of edge computing needs.

Although the data velocity and data volumes for planetary data may be orders of magnitude smaller than other domains, such as Earth science, they are large enough for the ML/AI/AA methods to be applicable and bring substantial potential value to planetary science. In some cases, planetary science data volumes have long since passed the point where traditional data analysis methods are practical. See also Section 4.6 of the PDS Roadmap Study (2017).

NASA recognizes the value of ML/AI/AA methods and is funding research and development projects for the application of these methods to planetary science. Although this review board is recommending some changes to the patterns of funding, we also recognize that NASA has funded most of those successful and valuable applications of ML/AI/AA noted in Finding 62.

As previously noted in this report (see Section 3.2.4) and in the PDS Roadmap Study (2017) Finding II: “There is a mismatch between the services and functions PDS is equipped to provide and the very high expectations of its users.” That mismatch extends to potential users and is particularly acute for ML/AI/AA applications. The fostering and development of a Planetary Data Ecosystem is an important part of addressing the unmet needs for ML/AI/AA applications and users.

Finding 63 Barriers to data access and data use have particularly strong impacts on the development and use of ML/AI/AA methods.

As discussed in this report and in the PDS Roadmap Study (2017), the many, and sometimes conflicting, expectations of the present PDS create barriers to access and use for some PDS users and applications. Some characteristics of the science of ML/AI/AA and of the communities of practice for ML/AI/AA methods result in these barriers having particularly strong effects for these applications and users.

The Machine Learning, Artificial Intelligence, and Advanced Analytics communities are interdisciplinary. They use common tools and infrastructure, and domain experts seek to apply and extend ML/AI/AA methods across scientific disciplines. Given the relatively small size of the ML/AI/AA community that works with planetary data, compatibility with these common tools and infrastructure is essential to fostering the use of ML/AI/AA methods in planetary science. However, as discussed in Sections 3.4 and 4.4 of the PDS Roadmap Study (2017), at least some PDS archival formats are difficult to use with analysis tools not specifically designed to work with these planetary formats.

ML/AI/AA work also requires understanding the linkages between data in a machine-actionable form, particularly linkages which describe provenance. Linkages are more than just connections between data sets. They also have attributes describing the nature of the linkage (some examples: one object can be part of, replace, contain, or be derived from another object). These linkages generally need to go deeper than the data set level (such as to the product or even variable level in PDS4). The globally unique nature of labels in PDS4 enables this kind of linkage, but we do not see mechanisms by which both the linkage and the attributes of the linkage can be expressed in PDS4.

ML/AI/AA methods often rely on higher-level and analysis-ready data products. As previously discussed (see Section 3.2.2), the mission and funding of the present PDS significantly limit the extent to which these higher level and analysis-ready data products can be archived. This, in turn, substantially limits the Findability and Accessibility of these data products, which has a particularly strong impact on the development and use of ML/AI/AA methods.

The FAIR data principles (see Section 2.1.4) specifically emphasize "machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data." This machine-actionability is particularly important for ML/AI/AA applications. We are not commenting on any specific deficiency in the present PDS or any other Ecosystem element, but rather pointing out the criticality of machine-actionability for the Ecosystem moving forward.

ML/AI/AA methods also depend heavily on access to data and metadata through automated means, preferably well-designed and robust Application Programming Interfaces (APIs). These APIs are critical to the ML/AI/AA community, and are also critical to the development of tools that will enhance the overall usability of the Ecosystem. PDS is developing these interfaces and prioritizes good documentation as part of that work. However, as noted in Finding 67, there are structural issues affecting software sustainability, which includes documentation and user support.

Finally, ML/AI/AA methods depend heavily on machine-readable data labels, such as image classifications. Many science communities have found crowdsourcing methods, including public participation in scientific research (also known as citizen science), effective for certain types of data labeling, which is necessary for ML/AI/AA applications. The Planetary Sciences Division, specifically, and the Science Mission Directorate, more generally, are actively engaged in multiple citizen science efforts, which this panel commends.

We reiterate here Recommendation 11, which is an-Ecosystem level item related to and motivated by the subject matter of this section: "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."

Recommendation 51 NASA should continue to foster the development of tools which translate from common planetary formats and standards into broadly used protocols, formats, and standards to enable the adoption of tools and methods in use by other science communities.

Recommendation 51 is substantively the same as the PDS Roadmap Study (2017) Finding IX: "There is a need for more translation programs that transform data from the PDS4 archive file formats to more usable analysis-ready formats."

Recommendation 52 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.

Recommendation 52 builds on Recommendation 51 by noting that higher-level and analysis-ready data products may not need the same level of curation and retention schedule as primary data. One path to addressing funding limitations for archival and delivery of higher order data products is to consider using levels of service and levels of archival rigor which depend on the nature of the data being archived.

Recommendation 53 NASA and the PDE should ensure that data linkage mechanisms and types are clearly documented with examples.

Recommendation 53 builds on Recommendation 47: “NASA should support and encourage expanded use of DOI-like identifiers for data, thereby connecting data at various levels of processing to assist users in locating the best version of a data set for their needs.” The IRB is making this particular recommendation due to ML/AI/AA methods depending heavily on machine-readable labels, as noted in the narrative following Finding 63.

Recommendation 54 NASA should find ways that the Ecosystem could include developer advocacy, particularly for the core PDS application program interfaces (APIs).

Recommendation 54 supports the ML/AI/AA community’s dependence on API access to data, as well as the software development needs discussed in Section 6.1.2. Developer advocacy is a form of user support, providing for channels of communication that help API users understand how to use those APIs in their applications and help API developers understand the ways that those APIs are, and are not, meeting user needs.

Recommendation 55 NASA should expand public participation in scientific research and other crowdsourcing methods as one strategy for providing data labeling essential to ML/AI/AA.

Finding 64 There are multiple structural issues that further impact the development and use of ML/AI/AA methods in planetary science.

There is at least anecdotal evidence that scientists with ML/AI/AA expertise are much more likely to be early-career professionals than in other aspects of planetary science. Structural issues that disproportionately affect early-career researchers, such as those discussed in Friesenhahn and Beaudry (2014), may therefore have a more profound effect on the development of ML/AI/AA methods.

Despite the demonstrated value of ML/AI/AA methods in planetary science, there is a perception among some members of the planetary science community that the nature of planetary missions, particularly the limitations of the Deep Space Network, profoundly limit the application of ML/AI/AA methods in this domain. Multiple ML/AI/AA practitioners (including IRB members) familiar with the application of these methods in multiple scientific domains hold the opinion that other domains have invested proportionally more of their resources in ML/AI/AA methods and that the funding opportunities for ML/AI/AA work in planetary science is not representative of the potential benefits of these methods in planetary science. These practitioners also observed that NASA ML/AI/AA funding opportunities in the planetary sciences are often scoped to specific missions, limiting the capability to extend the results of a project across missions or disciplines, and potentially excluding proposers who are not deep domain experts in the science of that specific mission or are not already part of a mission team.

Some ML/AI/AA practitioners expressed a view that their access to data and mission expertise outside that provided by the PDS was extremely important in their successful applications of these methods. This access included direct access to planetary data on servers, rather than having to work through the PDS interfaces and the bandwidth limitations of those interfaces, as well as getting answers to questions through personal connections with mission team personnel. Other practitioners expressed the perspective that technical and funding impediments to ML/AI/AA work with planetary data, combined with the wealth of opportunities in other domains, have caused them to deprioritize, or even abandon, their ML/AI/AA work in planetary sciences in favor of other domains, despite personal interests in planetary applications.

As the PDE continues to move forward in addressing opportunities to improve discoverability, usability, and interoperability of planetary data, NASA should consider creating funding opportunities that stress working across missions and/or creating, distributing, and maintaining higher level, analysis-ready data products.

The IRB notes that NASA is taking steps to address aspects of this finding, and we commend those efforts. NASA is actively evaluating the degree to which structural issues might disproportionately impact different groups of potential respondents, including early-career professionals. The Science Mission Directorate is also actively considering how to foster the development of ML/AI/AA technologies across the breadth of Directorate applications. The nature of the charge to this IRB is also indicative of the Planetary Science Division’s intent to foster the further development of ML/AI/AA methods in planetary science.

Recommendation 56 As they proceed with developing the Planetary Data Ecosystem, NASA should ensure that any Ecosystem assessment group considers the needs of current and potential ML/AI/AA users as part of their work.

Recommendation 57 NASA should also consider how the relatively nascent planetary ML/AI/AA user community might not be well-aligned with traditional missions, funding opportunities, and user groups and the impact that might have on potential respondents to funding calls.

While increasing funding for ML/AI/AA in planetary science is desirable, the IRB is recommending here that NASA consider how currently available funding might be more effectively allocated, given the observed barriers discussed here, including funding calls that work across missions and considering the lessons learned from other groups studying ways to foster ML/AI/AA across the Science Mission Directorate.

6.1.2. Multiple factors inhibit software development, sustainability, and adoption

Software is essential to the conduct of modern science. Software development can be an intimate part of the research process, where writing software to do something for the first time is an integral part of discovery and hypothesis testing. Other types of research software development are more engineering in nature, where the software objectives are relatively well defined, and the software must be reusable and robust. Software often operates in a hostile environment, whether in deep space or under attack by threat actors on the modern Internet.

In a research-driven environment such as planetary science, there is often no clear transition between the different aspects of research software development discussed above. Processes at a low Technology Readiness Level (TRL) generally will involve software that is more exploratory in nature, whereas processes at moderate to high TRLs will involve software where robustness and reliability are essential. Regardless of the TRL, ongoing maintenance is necessary for all software, both to meet evolving user needs and to ensure safe and secure operations in the modern dynamic technology landscape.

Finding 65 Current funding mechanisms and opportunities for software development and maintenance are insufficient and inhibit software development, sustainability, and adoption.

Current options for pursuing and obtaining funding under NASA R&A programs are insufficient and are not always well-suited for software development and maintenance efforts related to the PDE (and in general). Issues identified include the following:

- Modern planetary science requires software development to fully and efficiently use elements of the Planetary Data Ecosystem and there is insufficient funding to accomplish the necessary development. The only ROSES programs that explicitly fund software development are a portion of PDART (C.4), Support for Open Source Tools, Frameworks, and Libraries (F.7) and Supplemental Open Source Software Awards (F.8). They have not been resolicited under ROSES 2021.
- There is a shortage of funding opportunities to initiate the development of software tools in support of the PDE overall. This finding is similar to findings from the National Academies of Sciences, Engineering, and Medicine 2018 report on “Open Source Software Policy Options for NASA Earth and Space Sciences” (NASEM, 2018), and has not been addressed. The only

current program for creating new tools is PDART, which can fund only a small number of software proposals annually.

- Currently, there is no reliable mechanism by which software tools developed with NASA funds can be hosted, supported, or maintained past the expiration of the grants under which they were developed, beyond the nascent F.7 and F.8 appendices. Their value is therefore often lost or left unrealized.
- The cost (in time and funding) of preparing a typical ROSES grant proposal is essentially constant regardless of the scope of the work being proposed. Whether a proposal is requesting support for a “typical” 3-year effort by a moderately-sized team, or for a one-year effort by two people, the effective opportunity cost of creating that proposal is approximately the same. Therefore, proposers are incentivized to write proposals for larger awards even if smaller ones would be more appropriate. This issue is especially relevant for software development efforts, since many can be accomplished in relatively shorter timeframes and with smaller budgets compared to “typical” ROSES proposals. While such projects are possible to fund under current ROSES programs like PDART, the systemic incentives around proposal preparation and evaluation do not encourage that.
- Many of the individuals most qualified to initiate the development of software tools to support the PDE are in roles from which they are not typically allowed to lead “standard” research grants (e.g. technical staff, post-docs, etc.). Software tools development also remains undervalued in traditional academic career paths, so that many of the people who may apply for such awards are not strongly incentivized (or perhaps disincentivized) to do so. This topic is also covered in more depth in the National Academies of Sciences, Engineering, and Medicine 2018 report on “Open Source Software Policy Options for NASA Earth and Space Sciences” (NASEM, 2018).
- NASA has not historically used commercial contract mechanisms for funding software development relevant to the PDE, instead relying on R&A grant or cooperative agreement mechanisms. Contracts can offer some advantages, in some circumstances, such as by providing terms to enforce timely delivery of work product according to well-defined specifications, and incentivizing constrained project costs. Many organizations, including most commercial firms and many non-profit institutions, might also be especially well-qualified to develop software tools relevant to the PDE, but are not able to meet the administrative requirements of Code of Federal Regulations (OMB 2021) (but can meet the requirements of FAR more broadly).
- Current NASA solicitations for programs that explicitly support software development (C.4, F.7, and F.8 in ROSES 2021) do not provide sufficient detail to proposers or reviewers about how merit should be determined for a software development project relative to projects that primarily focus on scientific research. The criteria listed in the Guidebook for Proposers and ROSES 2021 are not necessarily appropriate or relevant for the evaluation of what are essentially software engineering projects.

Recommendation 58 NASA should increase the level of funding available to explicitly support software development, either via the existing ROSES programs or via the creation of new programs, and clarify its policies for evaluating funding proposals that do not include major components of hypothesis-based science.

Solicitations for proposals seeking this funding should take into account the diversity of types of software development that could be beneficial to the Planetary Data Ecosystem, including new data reduction and analysis tools, significant additions to existing software, tools for supporting the overall Ecosystem, conversion of software written in older and/or proprietary programming languages into more modern open-source code, and so on. NASA should review its current criteria for evaluating funding proposals that focus on software development and do not include major components of hypothesis-based science, ensure that those criteria are designed to evaluate the merits of such work as accurately as possible, and revise them if not. These optimized criteria should ideally be used to evaluate all software development projects within the Planetary Science Division or Science Mission Directorate.

Recommendation 59 NASA should establish a mechanism to support the preservation, support, and maintenance of software tools past the expiration of the grants under which they were developed.

There are many possible routes to this, including the following:

- "Support and maintenance" grants or contracts for terms of ~5 years for one or more software tools, at a total cost and level of effort appropriate to the nature and popularity of the software.
- The establishment (and financial support) of one or more organizations that would bear primary responsibility for the support and maintenance of PDE software tools, with terms of ~5 years. This might be a single organization or something similar to a new PDS node or a virtual institute.
- Given limited resources, the state of software "archiving" technology, and the nature of software systems, it may not be feasible to preserve all PSD-developed software functionality indefinitely. Resources should therefore be allocated based on NASA's priorities and community input, perhaps through the aforementioned PDE community-led assessment or analysis group.

Recommendation 60 NASA should consider providing options for funding software tools with the proposal requirements and total budget with the scale and scope of typical Guest Investigator or early-career programs.

Funding opportunities with these characteristics would provide a niche that is better suited in scope for small-scale software development, and will also incentivize proposers to request appropriately-sized awards by reducing the effort needed to prepare proposals asking for smaller awards. Such funding options could also provide opportunities for early career researchers to demonstrate their capacity for securing funds for software development projects, enhancing their future job prospects. In this way, such funding options could provide incentives for early career researchers to pursue software development work and reduce the perceived need to focus solely on traditional academic research at a time when their career prospects are still insecure.

Recommendation 61 NASA should consider, on a case-by-case basis, whether commercial contract bid mechanisms or grant proposal mechanisms would be more appropriate for efficiently filling certain critical Planetary Data Ecosystem software tool needs.

Strategic partnership programs, including Small Business Innovation Research and Small Business Technology Transfer (SBIR/STTR), are specific examples of programs where expanded usage could complement grant proposal mechanisms in meeting Ecosystem needs.

Finding 66 There is no uniform open-source software policy that applies across all Planetary Science Division elements.

This panel understands that there is an effort underway at the Science Mission Directorate level to implement an open-source software (OSS) policy, but it was not available at the time of writing this report.

Open-source software is encouraged in C.4 PDART, but it is not mentioned in other Planetary Sciences (ROSES Appendix C) calls, nearly all of which may involve software development.

Recommendation 62 Recognizing that “Software is data, but data is not software” (NASEM 2018 page 2), and in keeping with NASA’s open data policies, NASA should ensure that software developed by or for the Planetary Data Ecosystem is as open as possible and only as closed as necessary.

Recommendation 63 The Planetary Science Division should adopt a single, coherent, open source software policy that applies across all its activities. Ideally, this policy should be a consistent Science Mission Directorate policy. Given that portions of the Ecosystem are outside of NASA’s direct control, a single policy across the entire Ecosystem is likely not practical. However, it is appropriate for NASA to use its influence to achieve a high level of software policy consistency across the Ecosystem.

These activities include ROSES calls, missions, and direct-funding activities that fund the creation of software (specifically NASA class C, D, and E software). This policy should follow the best-practice recommendations of NASEM (2018) and Recommendations 2a, 2b, and 2c from the SMD Strategy for Data Management and Computing for Groundbreaking Science 2019-2024.

The creation and adoption of an updated open source software policy would require language changes in future ROSES calls, and mission announcements of opportunity. Existing missions could be brought into compliance by a variety of means (PLRA adjustment, project engineering handbook updates, senior review criteria, etc.). Finally, program officers that oversee directed funding programs (inter-agency agreements, the PDS, etc.) would need to understand the policy and how it would apply to software that those programs would produce. This policy should include an example Software Management Plan that can be used by ROSES proposals and missions in order to provide a guideline for what the policy expects.

NASA's open-source software policy, along with updated ROSES language and stated evaluation criteria, should reflect the critical nature of adequate documentation as part of proposed software development. Documentation enhances the accessibility and usability of software by others, and thus should be considered a key factor in the evaluation of proposed software development efforts. The ROSES 2021 call mentions documentation for proposed software development work in various places but does not specifically identify proposed documentation plans as a criterion for merit evaluation. To ensure that software produced under NASA-funded programs is understandable and usable by the widest audience possible, and to better guide both proposers and reviewers during the proposal process, future ROSES calls should include stronger statements (e.g., at least in Appendix C.1) about documentation being an essential part of software development and therefore explicitly subject to evaluation as part of the proposal review process.

Finding 67 There are inadequate training opportunities for planetary scientists in the use of software tools, as well as best practices for open-source software development, which limits software sustainability and adoption.

The combination of scientific needs unique to planetary science and the broader issue of a lack of interoperability with broadly used software tools leads to a large fraction of the tools in use being unique to planetary sciences. As identified in Finding 65, there are funding issues and other structural factors which inhibit software sustainability, and user training is one core element of sustainability.

Although some training exists for generic introductory level technical skills, there are very few opportunities for intermediate or advanced training in both software use and development relevant to planetary sciences. If scientists do not know how to use tools, the result is that those tools will not be used. Similarly, if scientists are not trained to create software using best practices associated with open-source software development, the result is that those tools will be of less utility to the broader scientific community. The issue of the availability of adequate training in software use is also relevant to issues of access and diversity, as training in more advanced tools may be effectively limited to those with connections or access to specific experts in the field, excluding those without such connections or access. This is similar to aspects of Finding 63, where the panel heard from ML/AI/AA practitioners who indicated that access to data and expertise outside that provided by the PDS was essential to their projects' success. It should be noted that this issue is particularly relevant in the realm of "big data" analysis as many producers of current and future large data sets explicitly intend those data sets to be of broad interest within the scientific community, meaning that large numbers of interested scientists may require training in any intermediate to advanced software developed to enable efficient analysis of those data.

Multiple members of the review panel have experience suggesting that one contributing factor to a scientific community having duplication and overlap of tools is a lack of awareness and understanding across the community of the capabilities of the individual tools. This board did not investigate this specific factor in the context of the Planetary Data Ecosystem, but we do conclude that it is likely an additional motivating factor for enhancing training opportunities.

As with software development and the structural issues associated with software sustainability, user training and technical documentation are generally done by people in roles from which they are not typically allowed to lead "standard" research grants (e.g. technical staff, post-docs, etc.). Documentation and user training also remains undervalued in traditional academic career paths, and even in software development communities, so that many of the people who are best qualified to do this work are not strongly incentivized (or perhaps disincentivized) to do so.

We echo the PDS Roadmap Study (2017) Finding VIII: "The PDS4 information model is well-documented at a highly technical level. However, there is a critical need for broader documentation and training for all levels of users." This training and documentation should also address data preparation from the perspective of reusability and interoperability, such as the Earth Science Data Systems Working Group (ESDSWG) Data Product Development Guide for Data Producers.

The panel recognizes that training is a difficult challenge. We did hear from subject matter experts responsible for training and education about the challenges of getting scientists to attend training when it is offered and to use the documentation that has been provided for tools.

Recommendation 64 NASA should seek to expand opportunities for intermediate to advanced technical training in topics related to accessing, using, and processing planetary data.

This expansion can involve expanding the kinds of funding mechanisms available for developing and conducting training and recognizing training as an essential element in software sustainability. This expansion can also involve clarifying the availability of funds to support this work through the NSPIRES opportunity "Topical Workshops, Symposia, and Conferences."

Recommendation 65 NASA should encourage continuing education of members of the planetary science community by making it clear that such costs are allowable on grants for all job categories.

7 Pathway Toward an Ideal State for the Planetary Data Ecosystem

7.1. Prioritizing

The high-level goal of this IRB is to help NASA to develop a seamlessly integrated Planetary Data Ecosystem that improves the planetary science community's access to, and use of, high-quality data. To convert our 65 recommendations into a pathway toward a more ideal state for the Ecosystem, these recommendations must be prioritized.

The Steering Group collected and individually ranked our recommendations by giving each a score of high, medium, or low. Discussion of the rankings yielded consensus on the top priorities. The following discussion presents only these top priorities as the IRB's recommended first steps toward a more ideal state for the Ecosystem. The full list of ranked recommendations has been provided to NASA via the spreadsheet used by the IRB Steering Group to determine the rankings.

The top-ranked recommendations fell naturally into three groups:

- Develop the Planetary Data Ecosystem.
- Address data preservation needs.
- Address barriers to data use and development.

All three groups of recommendations are of equally high priority; the order in which we present them should imply no relative ranking of the groups. However, within each group the recommendations are presented in order of their priority.

Priority order need not be the same as implementation order. All the recommendations listed below are considered high-priority, and some (the "low-hanging fruit") will be easier and faster to achieve than others. Therefore, the IRB envisions that a reasonable path forward might involve addressing high-priority and lower-priority recommendations from each group simultaneously.

We reiterate here what was emphasized in the Introduction (see Section 1.2): all three corners of the data stewardship triangle—data preservation, data discoverability, and data usability—must be seamlessly integrated to make progress toward an ideal state for the Planetary Data Ecosystem. We also encourage NASA to keep the IRB's core values in mind (see Section 1.3): the caretakers of the Ecosystem should strive toward it being open, collaborative, effective, practical, and compliant with FAIR data principles, without harm to the systems that are working now to enable NASA's achievements in scientific discovery.

7.1.1. Top Priority, Group 1: Develop the Ecosystem

The IRB expended considerable effort to understand the full scope of the Planetary Data Ecosystem (see Section 2). However, the final definition of its scope, responsibilities, governance, etc. (Recommendation 2 and Recommendation 3) will be up to NASA to explicitly clarify for each of its elements, including the PDS (Recommendation 13 and Recommendation 14).

The IRB (with the exception of one member) believes strongly that future development of the Ecosystem should be guided by a standing assessment or analysis group (Recommendation 4). Such a group is needed to help establish and maintain effective, on-going, multi-way communication among NASA, the broad planetary sciences community, and the elements of the Ecosystem. The group will ensure that the concept of the Ecosystem is communicated to the community, and that the needs of the community are clearly communicated to NASA. It is imperative that this group be broadly representative of all aspects of the user and developer community. This group is essential to help NASA continue to refine the full scope of the Ecosystem, to take the recommendations of this IRB and develop a long-term strategy, and to provide guidance as circumstances and knowledge evolve.

7.1.1.1. Recommendations to Develop the Planetary Data Ecosystem, in Priority Order

Recommendation 4: 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.

Recommendation 1: 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.

Recommendation 2: 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.

Recommendation 3: 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.

Recommendation 14: 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 prioritized 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.

7.1.2. Top Priority, Group 2: Address Preservation Needs

For nearly 40 years, reviews of NASA data governance have warned that irreplaceable planetary science data are being lost (see Section 3.1), and this Report is no exception. NASA must develop and implement a strategy to identify, prioritize, and preserve data at risk of loss.

Two of the preservation needs identified by the IRB were deemed so urgent that they should be addressed even before such a strategy is undertaken and completed. One is the preservation of planetary radar data (Recommendation 31). Some of these data have already been lost (e.g., the raw data for the discovery of ice at the poles of Mercury). The other is the need to preserve information and analysis related to returned samples (Recommendation 33 and Recommendation 34). These high-priority preservation needs are intricately interconnected with the high-priority need to establish guidelines for ways of archiving other than in the PDS, and thereby handle the ever-increasing data preservation needs referenced in Recommendation 28 and described in more detail via the examples in Section 4.

7.1.2.1. Recommendations to Address Data Preservation Needs in Priority Order

Recommendation 31: 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.

Recommendation 33: 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.

Recommendation 34: 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).

Recommendation 28: 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.

Recommendation 29: 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.

7.1.3. Top Priority, Group 3: Address Barriers to Use and Development

Group 3 includes a broad set of highly ranked recommendations geared to improving access, usability, and development across the Ecosystem. Broadly speaking, they illustrate the importance of effectively addressing the needs of the community (both users and developers) and improving their understanding of and access to Ecosystem elements.

7.1.3.1. Recommendations to Address Barriers to Use and Development in Priority Order

Recommendation 21: 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.

Recommendation 50: 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.

Recommendation 23: NASA should provide regular, accessible, and effective training programs should be provided 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.

Recommendation 64: NASA should seek to expand opportunities for intermediate to advanced technical training in topics related to accessing, using, and processing planetary data.

Recommendation 52: 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.

Recommendation 5: 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.

Recommendation 11: 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.

7.2. Forging a Pathway via a Case Study: Small-Body Science Data

The IRB would like to further illustrate a pathway toward an ideal state by describing a real-world example of what such an ideal state should encompass. From the many examples of data preservation needs discussed in Section 5, small-body science data stood out because it cuts across disciplines and research methods. Small-body science data touch upon a broad array of Planetary Data Ecosystem issues that apply to almost all areas of planetary sciences, including: uneven data discoverability; the difficulty of searching for data across numerous providers and repositories; metadata standards and quality; challenges of end-to-end data-to-result tracing; the need for better integration between software tools and data; and other issues unique to this research area. A cross-cutting working group on small-body research, initiated by the Utilization subcommittee and including members of other subcommittees, developed the following case study.

7.2.1. Data Discovery and Access Challenges

Small-body science has data discovery challenges, some of which are shared with other planetary science fields, but several of which are unique to the field:

- The non-sidereal motion of small solar system objects means that useful data can be obtained serendipitously by observers targeting other objects, performing wide-field surveys, or pointing beyond the solar system entirely but systematically identifying and cataloging such serendipitous observations is not straightforward.
- Geometric and orbital circumstances of small-body observations are often essential to their interpretation but are not typically readily available to data users, whether observations targeted the object of interest or not.
- 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 and approaches 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.

There is also a multitude of other data that are relevant to small-body science including derived physical properties, derived dynamical data, laboratory data, and numerical simulation and modeling data.

Data for specified small bodies or for objects meeting specified physical, dynamical, or observational criteria can be difficult to identify within the wide-ranging observational data ecosystem relevant for small body science. Tools like the PDS Small Body Ferret, JPL's Small Body Database, the Canadian Astronomical Data Centre's Solar System Object Image Search (SSOIS) tool, ESASky, and others exist to search for data on specific small bodies, but are typically limited in scope (i.e., in terms of the data sets that they are able to search) and also are not well integrated with each other. The data sets covered by these tools are typically identified on individual bases by the tool maintainers or the data submitters, meaning that these tools often have partially overlapping, but incomplete, coverage of the available data.

7.2.2. Toward a More Productive Small-Body Science Data Ecosystem

Engaging the full Planetary Data Ecosystem will be necessary to address the challenges identified above. It will require consideration of resources outside of NASA as part of the Planetary Data Ecosystem (e.g., other SMD divisions, domestic non-NASA entities like the National Science Foundation's NOIRLab, and international entities). In addition, data not currently stored on modern data storage media or held by facilities that currently do not maintain public archives must be recovered, digitized, and/or archived.

To successfully facilitate such science, NASA would need to:

- Evaluate the degree to which components of the Ecosystem relevant to small body science that are already managed under the NASA umbrella are integrated with each other, and consider ways to improve integration of those components. (Recommendation 5, Recommendation 11, Recommendation 16, Recommendation 18, Recommendation 26, and Recommendation 45)

  For instance, information from JPL's Small Body Database and HORIZONS ephemeris tool could be better integrated with PDS Small Body Node data and Minor Planet Center observation reports by allowing users to seamlessly access orbital and observational geometry parameters corresponding to those data or observations. Similarly, search results using JPL's Small Body Database Browser could include comprehensive lists of data available for particular objects across all PDS nodes.

- Support or incentivize the public archiving of analysis-ready data relevant to small-body science from observational facilities for which such data are not currently available, either because data are not publicly archived at all or only publicly archived in raw form. (Recommendation 28, Recommendation 32, and Recommendation 48)

  Notable examples of facilities that obtain data that are potentially relevant to small body science that currently lack public archives include Palomar Observatory, Las Cumbres Observatory, the Lowell Discovery Telescope, and TRAPPIST. Any observatory that routinely conducts imaging observations in principle have the potential to obtain serendipitous detections of small bodies, however, and so should be considered within the scope of this effort.

- Support efforts to improve the discoverability of data relevant to specified small bodies or small body observations meeting specified physical, dynamical, or observational criteria. (Recommendation 5, Recommendation 10, Recommendation 11, Recommendation 26, Recommendation 45, Recommendation 46)

  This could include enhancing integration between existing NASA search tools for small body data as described above, establishing partnerships with external institutions (e.g., CADC, ESA, ESO, etc.) with existing relevant tools, or creating new search tools or data portals (e.g., similar to the SIMBAD or NED tools that are widely used for astrophysics research) to facilitate the discovery of small body data meeting specific criteria.

- Encourage or incentivize the publication of pointing coordinate logs by observational facilities along with field-of-view and orientation information to facilitate the discoverability of serendipitous detections of small bodies. (Recommendation 28 and Recommendation 48)

  Searches of archival data for previously unrecognized serendipitous observations (i.e., “precoveries”) of small bodies can often be highly fruitful, enabling significant improvements to be made to their orbits and physical properties. Tools like SSOIS or ESASky can simultaneously search multiple data providers for such data by matching field-of-view information and observation times from observing logs to computed ephemerides of small bodies of interest. However, not all observing facilities make pointing and orientation information publicly available. Encouraging these metadata to be published, even for proprietary data which may not be public themselves, would expand the reach of small body search tools, create opportunities for new collaborations, and lead to new discoveries.

Although laying out a step-by-step strategy for reaching the ideal state is beyond the scope of this IRB, the use of case studies such as this one can help NASA to track the planetary science community’s progress along the path. The IRB’s top priority of establishing an advisory or assessment group for the Planetary Data Ecosystem would provide the community a prominent platform from which to highlight the development of Ecosystem components, allowing NASA to monitor the development of the Ecosystem and community members to be heard by NASA.

7.3. Endnote

During the course of our review one particular quote from IRB member, Kate Crombie, delivered in a meeting at the Institute of Space and Astronautical Science in March 2015 struck a chord with the IRB:

> *“The data gathered from missions is our National Treasure. It should be treated as such.”*

This sentiment was not only embraced by the IRB, but it was also quickly expanded upon:

> *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.*

A PDE IRB Charter and Subcommittee Charters

A.1. Planetary Data Ecosystem Independent Review Board Charter

A.1.1. Background

The NASA Planetary Science Division (PSD) is engaged in one of the oldest scientific pursuits: the observation and discovery of objects in our solar system. We undertake this enterprise in order to better understand the history of our solar system and the distribution of life within it. For decades, NASA's planetary science program has advanced the scientific understanding of the solar system in extraordinary ways, while pushing the limits of spacecraft and robotic engineering design and operations. The central component of that scientific pursuit is the data – from the research data that drives us to our next mission to the mission data received from the farthest reaches of the solar system. PSD is currently supporting an ad hoc, interconnected Planetary Data Ecosystem (PDE)[1] to take this data from creation to dissemination. However, the PDE is not yet effective and efficient at making the most of our planetary data and supporting our planetary science community. As such, PSD will take the first step necessary to develop a comprehensive PDE strategy – an independent review of the PDE in its current state to provide findings and prioritized, actionable recommendations that can be translated into an optimized PDE strategy.

A.1.2. Strategic Goals

For the future PDE, we expect a coordinated effort that meets the following strategic goals.

Strategic Goal 1: The PDE supports the data needs of the entire planetary science community, including planetary scientists and researchers; past, present, and future mission teams; educators and students; citizen scientists; media and the general public, where appropriate.

Strategic Goal 2: The PDE fills the prioritized, required niche space, which is supported by stable PSD funding sources, and includes element redundancies only where necessary to support the needs of the planetary science community.

Strategic Goal 3: The PDE adapts to meet the evolving data and computing environments (e.g. cloud computing, new formats, ecosystem cybersecurity) and needs of the planetary science community. As such, the PDE will have pathways to onramp and offramp elements to address

---

[1] The Planetary Data Ecosystem (PDE) is defined as 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, primarily those funded by NASA. The PDE includes archives, tools, programs, projects, pipelines, and stakeholder groups, but that is not an exhaustive list. This definition is for the purposes of this document and should be updated by the PDE IRB to more accurately reflect the current state of the PDE.

changes in the data environment and a transparent process for prioritizing competing community needs.

Strategic Goal 4: The PDE communicates across elements and has policies to facilitate effective work across a fully coordinated system.

The PDE IRB shall provide prioritized, actionable recommendations that would help PSD meet these strategic goals.

A.1.3. Establishment, Authority, and Study Management

The PSD Director establishes the PDE Independent Review Board (IRB), hereafter the “PDE IRB” As such, the PDE IRB will report to the PSD Director.

This PDE IRB shall be organized by Cornell Technical Services (CTS), with additional support from NASA Research and Educational Support Services (NRESS). Members of the PDE IRB will be compensated for their participation and will, upon request and when possible, be allowed travel and per diem expenses as authorized and funded by PSD. The PSD Director will ensure the necessary support for the PDE IRB, including appointing NASA personnel as a Review Manager and ex officio members, as needed. The PDE IRB Chair and the Review Manager will support all activities of the PDE IRB and coordinate production and ensure the quality of review deliverables. The Review Manager will ensure that the information needs of the review members are met. The non-consensus final report will be verbally presented to the PSD Director and other NASA stakeholders, followed by the provision of a non-consensus final written report.

A.1.4. Purpose and Scope

The PDE IRB will conduct a wholistic review of PSD’s PDE with the goal of defining the full environment, identifying missing or overly redundant elements, and providing findings and prioritized, actionable recommendations for PSD’s long-term planning in support of the PDE. The scope of the PDE IRB, outlined in the Key Questions below, is broad and their deliberations will cover the entire data lifecycle, including initial processing, long-term archiving, scientific searches and utilization, ecosystem cybersecurity, ecosystem governance, and public dissemination. The scope of the PDE IRB will not include review of specific activities or projects within the PDE, as that is completed by a separate review process (e.g. Senior Review, peer review process). A successful PDE IRB will provide prioritized, actionable recommendations, that if implemented by PSD, would significantly advance the PDE towards a future state consistent with the Strategic Goals.

Key Questions:

1. Ideal State: What PDE elements are necessary to take planetary mission data from downlink to research outcomes (e.g. a successful data lifecycle)?

2. Current State: What are the current PDE elements and what does each element provide in terms of the data lifecycle?
3. Comparative State: What PDE elements are missing and where are there PDE element redundancies?
4. Future State: What is the PDE IRB recommended pathway to evolve the PDE from its current state to an ideal state?

A.1.5. Organization and Duties

The PDE IRB will be organized into Sub-Committees based on functional aspects of the PDE:

1. Archiving (storing data, formatting, metadata requirements, documentation)
2. Searching (discoverability of data, open access, translation)
3. Utilization (tools for analyzing data, open software, software development)
4. Mining & Automation (searching and using large datasets, high-end computing, cloud computing)
5. Inter-relational (relationships among the ecosystem elements, ecosystem cybersecurity, ecosystem structure, ecosystem governance)

The elements of the PDE listed in Appendix A may or may not be a comprehensive list. Additionally, the Sub-Committee organization is open for negotiation by the PDE IRB Chair. The Members of the Sub-Committees will conduct research and fact-finding as required to fulfill their duties, which will be more specifically enumerated in a kick-off presentation provided by the NASA ex officio member(s) of the PDE IRB. Each Sub-Committee will be led by a Co-Chair, and the Sub-Committees will report their findings to the full PDE IRB for its consideration.

A.1.6. Membership

The PDE IRB will consist of Members, selected to ensure a balanced representation across the topic areas of the Sub-Committees. The PDE IRB will be composed of Members from academia, industry, Government, and the general public and will span the necessary expertise areas of governance, project management, science, engineering, and user base. The PDE IRB Members will have, collectively, the expertise and experience required to provide a wholistic and accurate assessment of the PDE.

Recommendations for the Chair of the IRB, Co-Chairs of the Sub-Committees, and Members will be provided by Lori Glaze, the Director of PSD, based on input from NASA Headquarters personnel and self-nominations. The Members will be assigned to the Sub-Committees by CTS, in consultation with the Chair, Co-Chairs, and NASA ex officio member(s), and will participate only in the work of their appointed Sub-Committees, unless agreed upon otherwise.

A.1.7. Administrative Provisions and Schedule

The PDE IRB will meet in its entirety at least three times, either in-person or virtually, as appropriate and necessary to meet its responsibilities; the Sub-Committees will meet in addition to this as needed. We anticipate engagement at the Sub-Committee level on a weekly or biweekly basis over a period of about 3 months and estimate the level of effort to be 3 to 4 hours per week plus the possibility of travel. The PDE IRB meetings will be open to the public, but the Sub-Committee meetings will be closed.

The PDE IRB will draw on the expertise of its Members along with formal input solicited from the scientific community (white papers and invited discussions/presentations) and existing documentation to formulate its findings. The NASA ex officio member(s) will provide the Sub-Committees with a curated document library. A call for one- to two-page white papers will be announced to the community to provide additional input to the PDE IRB. It is anticipated that the majority of white papers will be received prior to the first meeting of the PDE IRB, but it will be possible for white papers to be submitted later in the process. Lastly, the Sub-Committees may request a presentation or discussion on a particular topic to be facilitated by the appropriate project personnel. The NASA ex officio member(s) will support these efforts as possible.

Tentative Schedule:

Week	
Pre-Week 1	Select and appoint PDE IRB Members Publish PDE Request for Information (RFI) to solicit responses from community
1	PDE IRB Organizational Telecon (1-hour); review background materials for all Sub-Committees Sub-Committee Break-out Session (2-hours); NASA ex officio member(s) presents additional guidance on Sub-Committee focus
2	Sub-Committee Meetings #1 – Identify additional presentations/discussions needed
3	Sub-Committee Meetings #2
4	PDE IRB Half-day Meeting
5	Sub-Committee Meetings #3 – Additional presentation/discussion
6	Sub-Committee Meetings #4
7	Sub-Committee Meetings #5
8	PDE IRB Half-Day Meeting
9	Sub-Committee Meetings #6 – Additional presentation/discussion
10	Sub-Committee Meetings #7 - Writing assignments/discussion
11	Sub-Committee Meetings #8 – Writing assignments/discussion
12	PDE IRB All-day Meeting
13	---

14	Develop and discuss draft findings for full PDE IRB report; decide on writing assignments
15	Draft any final questions for further discussion
16	Work on writing assignments and internal review; submit draft report to NASA and NASA review of draft report
17	Review NASA comments on draft report; close out remaining questions and revise draft report; and NASA reviews revised draft report and submits final comments
18	Complete draft report
19	Brief non-consensus final report to PSD Director and other NASA stakeholders in Washington, DC or virtually
20	Prepare non-consensus final report; deliver to PSD Director

A.1.8. Deliverables and Termination

The PDE IRB will provide a non-consensus report with observations, findings, concerns, and prioritized, actionable recommendations consistent with the above Purpose and Scope to the PSD Director. The PDE IRB will review the reports of its Sub-Committees and include them, if adopted, in its final report to NASA. Additionally, the PDE IRB will provide a presentation to the PSD Director and other NASA stakeholders summarizing the review results. The PDE IRB shall deliver its final report and presentation by 31 March 2021.

A.2. PDE IRB Charter Appendix: Subcommittee Details

A.2.1. Archiving

(storing data, formatting, metadata requirements, documentation, redundancies and gaps)

Summary of Sub-Committee Objectives:

The Archiving Sub-Committee will analyze and assess the current state of the PDE archiving processes and archive locations to provide findings and recommendations on how PSD data is effectively being archived. This Sub-Committee will focus on whether data are appropriately being prepared for archiving, if the right resources are available to take data from its source to an archive, and whether an appropriate archive is available to the community. This Sub-Committee will tackle topics related to:

- Standard formatting (Standard data types used in the community?)
    - Are the current data standards working for the community that works with the data?
    - Do all the archives and databases within the PDE work effectively together?
- Documentation/metadata (Sufficient to enable going from zero to functional?)
- Submission tools (Sufficiently enforce format/metadata standards?)
- Improvements/upgrades (Upgradable as processing, etc., improves?)
- Long-term curation (Long-term maintenance of archived data, models, etc.)
- Redundancies and Gaps (multiple archives, centralized resource, etc.)
    - What PDE archives and databases include the same data and, if so, what value does that add to the PDE?
    - Is there a PDE element for all the data the community is interested in accessing?

PDE elements included:

- PDS archives
- Discipline node (DN) data holdings: Atmospheres, Cartography & Imaging Sciences, Geosciences, Planetary Plasma Interactions, Ring Moon Systems, Small Bodies; also NAIF support node
- DN Data Node holdings: THEMIS, HiRISE, LROC
- Registry Catalog
- NSSDCA
- Spectral data archives and collections: HITRAN/Harvard, RELAB/Brown Univ. (coming to PDS Geo), TIR/ASU, VIS-NIR/USGS, CRISM spectral library (PDS Geo)
- Sample archives and databases
  NOTE: The initial PDE IRB will NOT review sample curation of returned samples, meteorites, IDPs, or those collected at terrestrial analogs. While it is understood that data on these samples may be contained in some elements of the PDE, it has been determined that in order to prevent too much expansion of the scope of the PDE IRB, and the expertise required, this subject area will be excluded. It is PSD's intention that the issues in this area (such as the lack of an archiving system for terrestrial analog data) will be assessed later and fully, intentionally,

integrated into the PDE. As such, these elements will NOT be prioritized for review by the PDE IRB:
    - Astromaterials Acquisition and Curation Office JSC
    - Meteorite samples
    - Asteroid samples: Hayabusa/Itokawa samples, OSIRIS-REx/Bennu samples
    - Genesis, Stardust/Wild 2, Cosmic Dust, Microparticles, etc.
- However, these elements will be prioritized for review by the PDE IRB:
    - JSC Curation online catalogs, such as:
        - Lunar sample catalog: Apollo rock sample image archive, Apollo Samples and Photo catalog, Lunar Sample Compendium
        - Antarctic Meteorite Database
    - AstroDB (inc Moon DC; format derived from/related to PetDB, Columbia Univ.), MoonDB/Astromat
- Data archive tools
    - PDS/PDS4: Generate, OLAF, PLAID/APPS, Validate, Transform, PDS4 JParser, GDAL, MakeLabels, Local Data Dictionary Tool (LDDTool), Docgen – create PDS4 labels from PDS3 labels – PPI, Mimic – stream data, synchronize servers - PPI
    - HAPI (Heliophysics API) – stream time series – Heliophysics Division

A.2.2. Searching

(discoverability of data, open access, translation)

Summary of Sub-Committee Objectives:

- The Searching Subcommittee will analyze and assess how easy it is for a community user to navigate the PDE and whether there are improvements that could be made to facilitate this navigation. This Subcommittee will provide findings and recommendations on how open the PDE and its elements are to a variety of user types (e.g. established researchers, citizen scientists). This Sub-Committee will tackle topics related to:
- Open Access (Effective open data policy? User-friendly access to datasets?)
    - Does PSD have an effective open data policy that is implemented across PDE elements? How would compliance with such a policy be enforced?
- Discoverability (Easy to find datasets? Search by metadata? Synthesized?)
- Usability (Datasets are “science-ready” / “analysis-ready” for end users?)
- Translation (Enhanced data usability by translating into other formats?)
    - Does the PDE enhance data usability by providing resources for users to train themselves or others to become more sophisticated data users?

PDE elements included:

- Archives (see Archiving Sub-Committee elements)
- Tools (see Archiving and Utilization Sub-Committees elements)
- General Public Data Access
    - Planetary Photojournal

A.2.3. Utilization

(tools for analyzing data, skill development, open software, software development)

Summary of Sub-Committee Objectives:

The Utilization Sub-Committee will assess and analyze whether the infrastructure of the PDE is effective in allowing the community of users to work with the data and if the appropriate resources are available to facilitate this work. Additionally, this Sub-Committee will explore whether the program infrastructure to support tool and software development is sufficient to keep the PDE evolving and stable. This Sub-Committee will tackle topics related to:

- Open Software (Effective open software policy?)
    - What definitions and language are appropriate to differentiate needs and policies as they relate to software?
- Discoverability (Easy to find/use software tools?)
- Development (Support for software development, esp. early career?)
    - Are citizen scientists supported by our development of tools and software to utilize the PDE?
    - How do elements developed in the PDART program move to be key elements of the PDE with long-term, stable resource support?
- Data Skill Development (Skills for providing foundational products?)

PDE elements included:

- Community data visualization, search, and access/retrieval tools
    - PDS
        - Analyst's Notebook (Geo), Atlas (CIS-JPL), Annex (CIS-USGS), Ferret (SBN), Marsviewer (CIS-JPL), Small Bodies Image Browser (SBN), ODEs (Geo), OPUS (RMS), Viewmaster (RMS), UPC,PILOT (CIS-USGS), Map-a-Planet (CIS-USGS), Photojournal (CIS-JPL), PDS4 Viewer (SBN), NASAView (EN), Cosmographia (NAIF), Spectral Library (Geo), Aspera high-speed data transfer (Geo), WebGeoCalc (NAIF), CATCH (SBN), MPC live DB distribution (SBN), MPChecker (SBN - MPC), MPC DB query tool (SBN - MPC), Splash – time series data analysis and display – PPI, PNG Walks PPI (Iowa), VISTA PPI
    - USGS
        - Planetary Nomenclature (on behalf of IAU); Working groups for planetary bodies, coord syst, etc.
        - Geologic Mapping & GIS: MRCTR and GIS data and tools; apatial data access initiatives; community training events
        - IAA Spatial Data Infrastructure: Astropedia; Geospatial data metadata, cataloguing, access initiatives; ISIS3 planetary cartography software; Geodesy program (Coordinate systems, reference frames, IAU international coordination and working groups); "Cartography" program and data products; community training events
    - JMars/Moon, etc.
        - Geospatial data access, tools, visualizations, processing

        - Community training events
    - Treks
        - Geospatial data access, tools, visualizations
        - Community training events
    - Quickmap, ACT (Erick Malaret)
    - RPIFs
        - Hardcopy, unique data collections, books, maps, etc.
        - Digital data access, tools
        - Community outreach and training events
    - Small Body Mapping Tool
    - Planetary Data search tools for/within Astrophysics Division:
        - ADS (often used to find PSD products, including PDS holdings)
        - IRSA NEOWISE image viewer & photometry search (IPAC, Caltech)
    - Autoplot – University of Iowa
    - Topcat – Virtual Observatory Alliance
- Community data processing tools
    - Ames Stereo Pipeline
    - GDAL
    - MRCTR GIS lab and tools
    - Projection on the Web (POW) tool at USGS
    - Mission tools (sometimes shared within teams only): HiView, HiRISE, Univ AZ; LunaServ, LROC, ASU
    - Quickmap MESSENGER/LROC/M3, ACT
    - Titan Swatch Viewer (TSV), Cassini team
    - Community toolkits and collaborations: OpenPlanetary, PlanetaryPy, SpicePy, etc.
    - PDS: SPICE Toolkit, WebGeocalc (NAIF), Others
    - AI4Mars
- Planetary Spatial Data Infrastructure (PSDI) elements
    - Foundational products
        - Missions
        - Community derived products
        - Geodesy program
            - GSFC elements
            - Missions
    - Policies
    - Standards
        - PDS
        - FGDC
        - ISO
        - IPDA collaborations, tools, services
        - International coordination with IAU, national space agencies such as ESA/PSA, JAXA,ISRO, CNRS, IVOA etc.
    - Working groups, collaborations
- NASA Derived Data Products, Software
    - PDART
    - DAPs
    - Other

- NASA GitHub
- NASA Planetary Github
- PDS tools and databases
- USGS Astrogeology ISIS3
- NASA software
    - From missions
    - From DAPs

A.2.4. Mining & Automation

(searching and using large datasets, high-end computing, cloud computing, ML/AI)

Summary of Sub-Committee Objectives:

The Mining and Automation Sub-Committee will analyze and assess how emerging computing techniques, resources, and services support the PDE and PDE users. In particular, this Sub-Committee will focus on current and future best practices for a data ecosystem that should be brought to fruition for the PDE to meet current and emerging user needs and improve the PDE's effectiveness in serving the community. This Sub-Committee will tackle topics related to:

- Access / Cloud Computing (Commercial services meet needs? Plans for future?)
    - What would be the advantages and disadvantages of moving PDE elements to the cloud environment?
- Analysis / HEC (Sufficient resources provided to community? Reasonable application process? Resources up-to-date?)
    - Do the current HEC resources meet the computing and processing needs of the community?
- Automation / ML (How is this being used/coordinated in PSD? Plans for future? What can/should the PDE do that best enables users to use PDE with AI and ML methods?)
    - Example. Are the current investments in automation and streamlining sufficient to see maximal gains?

PDE elements included:

- High-end computing (HEC) facilities
    - HECC/NAS, ARC: Data Portal, ECCO, HELIO, QUAIL
    - NCCS, GSFC
    - Others?
- Cloud computing resources (NASA, commercial)
    - Amazon AWS
    - Microsoft Azure
    - Nebula IaaS/Paas/SaaS/DBaaS, ARC
    - Google Cloud
- PDS data services
    - PDS API
    - Data in Cloud?
- AI and machine Learning

  - JPL Machine Learning and Instrument Autonomy Group
  - JPL Artificial Intelligence Group
  - AI4Mars

A.2.5. Inter-relational

(relationships among the ecosystem elements)

Summary of Sub-Committee Objectives:

The Inter-relational Sub-Committee will analyze and assess the stakeholder and management elements of the PDE. This Sub-Committee will provide findings and recommendations on possible ways to make the PDE more cohesive and clearly connected. This Sub-Committee will provide insight into potential advantages and disadvantages of management and communications structures and identify gaps that should be filled in order to boost the collaborative nature of the PDE. This Sub-Committee will tackle topics related to:

- Ecosystem governance
- Ecosystem cybersecurity
- Discoverability of data (SC2) relies on searches/filters based on metadata (SC1)
- Software/models (SC3) requires datasets (SC2), should these be stored in the same place (SC1)?
- Community outreach
  - How can we discover how well the community is being served by the PDE?

PDE elements included:

- Project offices and program management: PDS, NSSDCA, PDCO (MPC, Harvard CfA/SAO; CNEOs, JPL; IAWN, UMD) AMMOS, JPL
- Mission-critical/Mission-enabling activities: JPL HORIZONS system, JPL NAIF/SPICE, Telescopes (incl radar? Goldstone? DSNs? TDRS?), Planetary Spatial Data Infrastructures, community outreach and training efforts / skill development (see also panel 1), training by PDS for PDS, training by PDS for community (including SPICE training classes), outreach events at national meetings, meetings (domestic and international)
- Planetary Data Workshops
- PSIDA
- Mission team events: by and for teams, for public and data users, tools and tutorials (including SPICE tutorials and self-training packages)
- Stakeholders: NASA HQ, Other NASA divisions, NSF, NASA Missions, International missions, Individual data providers, PDS, planetary data users, PDS Data and Archive Working Groups, MAPSIT, CAPTEM, IPDA, Cartography Working Groups, general public (educators, students, citizen scientists)
- Relationships with and among stakeholders: communications, support, feedback from user groups, surveys, etc.

A.3. Archiving Subcommittee Charter

Archiving covers data storage, formatting, metadata requirements, documentation, redundancies, and gaps, as well as data standards development.

A.3.1. Summary of Archiving Sub-Committee Objectives:

The Archiving Sub-Committee will analyze and assess the current state of the Planetary Data Ecosystem (PDE) archiving processes and archive locations in order to reveal findings on how Planetary Science Division (PSD) data are currently being archived, and provide recommendations for improved efficacy. This Sub-Committee will focus on whether data are appropriately being prepared for archiving, if the right resources are available to take data from its source to an archive, and whether all PSD data have an appropriate archive that is available to the community. This Sub-Committee will tackle topics related to:

- Standard formatting (Standard data types used in the community?)
    - Are the current data standards working for the community that works with the data?
    - Do all the archives, repositories, and databases within the PDE work effectively for the community?
    - Do some archives, repositories, and databases need to work together? Do work together effectively?
- Documentation/metadata: Are the metadata sufficient to enable going from zero (pre-acquisition of data) to functional (searching, use, analysis, and contextualization of the data)?
- Do submission policies tools sufficiently enforce format/metadata standards?
- Do any PDE data archives, repositories, and databases require Improvements or upgrades now or in the near future?
- Long-term curation (Long-term maintenance of archived data, models, etc.)
- Redundancies and Gaps (multiple archives, centralized resource, etc.)
    - What PDE archives and databases include the same data? If overlap exists and, if so, what value does that add to the PDE?
    - Is there a PDE element for all the data the community is interested in accessing?
    - How do PSD archives interact with/overlap with Astrophysics and other non-PSD archives or repositories?
- Do the archives lend themselves to being used (searched, mined, etc) in the future?

A.4. Searching Subcommittee Charter

A.4.1. Summary

The Searching Subcommittee will assess how well the Planetary Data Ecosystem serves its user communities by making its holdings easy to discover, navigate, and browse, and to search for and retrieve data. The Subcommittee will consider the quality of the users' experiences of identifying and selecting products of interest and for retrieving them or interacting with them on remote servers. The subcommittee will provide findings and recommendations for improving the search experience for different user communities.

A.4.2. Objectives

- Define the subcommittee's scope:
    - Review the existing list of PDE elements with the goal of developing a more comprehensive list through the process of refining, adding, editing and/or deleting entries.
    - Develop a list of potential search personas that represent a spectrum of users, their search goals, and how they would use existing tools to discover their desired data.
    - Identify the challenges these potential users encounter when searching for their desired data.
    - Identify pathway(s) by which users can successfully search, select, and retrieve their desired data
- Assess the quality of the user search experience:
    - Match the PDE list with existing search tools and identify gaps in tools and/or pathways to locate desired data.
    - Identify areas where existing tools either do not exist or where the user needs to have intimate community knowledge or prior knowledge of data location.
    - Identify barriers to data discovery via examination of individual persona attempts at data discovery
    - Identify redundancies in locating or discovering data and assess their utility or potential to hinder, i.e. beneficial (e.g. provide more than one type of path to the data to meet different user needs), neutral (redundant but with little additional overhead cost), or negative (redundant and duplicative of effort)
    - Assess the overall quality of the user's experience.
- Develop findings and recommendations:
    - Identify PDE data holdings that lack usable search tools, or provide poor user experiences
    - Identify PDE data holdings that provide positive user experiences (e.g. types of search goals that can be satisfied) and can serve as models or guides for future development.
    - Identify data that is not currently available in the PDE that, if included, would improve data discovery and user experiences.
    - Recommend actions for improved data discovery and query to locate their desired data.
    - Prioritize recommendations.

A.5. Mining and Automation Subcommittee Charter

A.5.1. Objectives

The Mining and Automation Sub-Committee will analyze and assess how emerging techniques, resources, and services support the PDE. In particular, this Sub-Committee will focus on current and future best practices for a data ecosystem that should be brought to fruition for the PDE to stay relevant and effective at serving the community.

A.5.2. Scope

The scope includes searching and using large datasets, high-end computing (HEC), cloud computing, Machine Learning (ML) & Artificial Intelligence (AI)

A.5.3. Key (strategic) questions we need to address

- How can NASA better enable PDE users to advance science through data mining, machine learning, and other analytics methods?
- How can NASA better enable PDE users and data providers to use automation, cloud computing, and other advanced computing methods with PDE data, particularly in support of analytics and making their work more efficient/effective?
- How can NASA use mining, machine learning, analytics, automation, cloud computing, and other advanced/emerging computing methods to improve the PDE's ability to meet user needs (operational excellence)?
- For this PDE today, in the domain of mining and automation, what needs to change, what do users need more of, what do users need less of, and what absolutely must not change?
- How do the answers to these questions vary across the spectrum of users?

A.5.4. Example questions to feed strategic questions

- Access / Cloud Computing
    - Would commercial services be able to meet PDE needs?
    - What would be the (dis)advantages of moving PDE elements to the cloud environment?
- Analysis / HEC
    - Are sufficient, up-to-date resources being provided to community?
    - Is the application process for these resources reasonable?
- Automation / ML & AI
    - How is automation & ML being used/coordinated in PSD? What are the plans for the future?
    - Are the current investments in automation and streamlining sufficient to see significant gains?
- Relationship to Other Sub-Committees
    - How do policies for archiving of software/models and their output affect the need for cloud computing and HEC resources?

    - How does the open source software model impact the future of ML, AI and other modeling efforts?

A.6. Utilization Subcommittee Charter

A.6.1. Summary

The Utilization Subcommittee will assess and analyze whether the infrastructure of the current PDE allows the user community to work effectively with available planetary data and if appropriate resources are available to facilitate this work. Additionally, this subcommittee will explore whether the program infrastructure to support tool and software development is sufficient to sustainably address PDE needs.

A.6.2. Key questions to address

- Ideal State: What PDE elements are necessary?
- Current State: What are the current PDE elements?
- Comparative State: What PDE elements are missing?
- Future State: What is the recommended pathway?

A.6.3. Scope

The Utilization Sub-Committee will assess and analyze whether the infrastructure of the current PDE is effective in allowing the community of users to work with available planetary data and if appropriate resources are available to facilitate this work.

Additionally, this Sub-Committee will explore whether the program infrastructure to support tool and software development is sufficient to keep the PDE evolving and stable.

This Sub-Committee will tackle topics related to:

- Open Software (Effective open software policy?)
    - What definitions and language are appropriate to differentiate needs and policies as they relate to software?
- Discoverability (Easy to find/use software tools?)
- Development (Support for software development, esp. early career?)
    - Are all user groups supported by our development of tools and software to utilize the PDE?
    - How do tool and software elements developed in the PDART program move to be key elements of the PDE with long-term, stable resource support?
    - What other pathways exist for supporting tool development and long-term support? Are they sufficient?
- Data Skill Development (Skills for providing foundational products?)
- Is our software development community healthy? Too big/too small? The right mix of skills?

A.6.4. Objectives - short and long term

- (short) Determine accurately the subcommittee's scope:

  - Define concepts/topics that would help the classification of the PDE tools.
  - Evaluate the needs of the tools as a function of personas that are using the tools (expert, beginner, etc..).
  - Evaluate how tools generated through various funding sources (e.g, mission, PDART, etc..) are distributed and available to the community.
  - Define the overlap with the searching and archiving sub-committees (and potentially others as well), and propose a common plan of work if necessary
- (medium) Assess the ability for the user to find, get access, and use tools for scientific data analysis:
  - How do users find tools, is there one single point of entry, or numerous ways to access tools for planetary science? Are tools registered somewhere? How is the PDS registry helping for this?
  - What licenses are required to use the tools ? Are a majority of the tools open access? Shall the open access policy be increased?
  - What is the level of redundancy between the tools? Are fundamental domains lacking tools? Is redundant work being done recreating existing, but not accessible, tools?
  - What is the level of maintenance of the tools? Are critical tools not well-supported for the long term?
  - How can the PDE be optimized with regards to tools? How can barriers of locating and using be lowered?
  - Are they any categories of personas that are significantly left out?
- (medium) Assess the support, guidance, and requirements provided to developers to make their tools accessible and usable by users:
  - Is there any “standard” that helps to streamline the usability of tools? Similar to PDS4, are rules available to guideline the development and dissemination of tools?
  - What is the level of development of new tools with respect to the new missions, new programming languages, PDS4 format, etc..?
- (long) Develop findings and recommendations:
  - Summarize the key findings of the sub-committee regarding
  - Access to tools
  - Usability of tools for the entire community
  - Left-out users, or topics not properly covered by tools
  - Propose recommendations to NASA regarding the key findings
  - How can NASA support long-term management, dissemination, and support of developed tools
  - Provide a prioritized list of the recommendations

A.7. Inter-relational Subcommittee Charter

(relationships among the ecosystem elements)

A.7.1. Summary of Sub-Committee Objectives

The Inter-relational Sub-Committee (IRSC) will analyze and assess the various stakeholder elements of the PDE (management, users, agency leadership, etc.), and will provide findings and recommendations on possible ways to make the PDE more cohesive and clearly connected. This Sub-Committee will provide insight into potential advantages and disadvantages of management and communications structures, and will identify gaps that should be filled in order to boost the collaborative nature of the PDE. We will examine and understand these elements and the relationships among them, and recommend improvements if necessary. These insights will be gained by examining topics related to:

- PDE data access and security
- PDE infrastructure, architecture (structure, metadata, taxonomies, other related information science elements) and governance policies
- Inventory of current PDE elements, relationships among these elements, and identifying missing and redundant elements.
- Effectiveness of the current PDE elements to enable and facilitate scientific research and innovation
- Searches/filters based on metadata and how these apply to data discoverability
- Storage of software and dependent databases
- Community outreach and effectiveness of awareness efforts
- Identification of stakeholders and relationships between relevant groups with, and relationships among stakeholders
- Recommended pathways to evolve the PDE from its current state to an ideal state

A.7.2. PDE elements included

- Project offices and program management
- Mission-critical/Mission-enabling activities
- Planetary Data educational components
- IPDA and other data-driven international agreements

B IRB Structure and Membership

Planetary Data Ecosystem
Independent Review Board
Team Membership

NASA Review Manager
Rebecca McCauley Rench

Chair
Melissa McGrath

CTS Task Manager
Autumn Manecke

NRESS Support
Susan Borden
Latessa Tuck
Jason Silveira

Archiving Subcommittee
Co-Chair
Moses Milazzo
Executive Secretary
Sarah Black
NASA Ex-Officio Members
Meagan Thompson
David Smith
Members
Kate Crombie
Martha Maiden
Joe Masiero
Flora Paganelli
Corrine Rojas
Priyanka Sharma
Marshall Tabetah

Searching Subcommittee
Co-Chair
Emily Lakdawalla
Executive Secretary
Katya Gilbo
NASA Ex-Officio Member
Lindsay Hays
Members
Julie Castillo-Rogez
Vandana Desai
Mary Ann Esfandiari
Laszlo Kestay
Matthew Miller
Lynn Neakrase

Utilization Subcommittee
Co-Chair
Sebastien Besse
Executive Secretary
Don Hood
NASA Ex-Officio Member
Sarah Noble
Members
Alyssa Bailey
Ross Beyer
Henry Hsieh
David Mayer
Chase Million
Danielle Wyrick

Mining & Automation Subcommittee
Co-Chair
Bruce Wilson
Executive Secretary
Hannah Dattilo
NASA Ex-Officio Members
Megan Ansdell
Kenneth Hansen
Members
Abigail Azari
Benjamin Burnett
Amitabha Ghosh
Jason Laura
Jennifer Shin
Orenthal J. Tucker
Alexandria Ware

Inter-Relational Subcommittee
Co-Chair
Caroline Coward
Executive Secretary
Blake Mendoza
NASA Ex-Officio Members
Steven Crawford
Lucas Paganini
Members
Reta Beebe
Shawn Brooks
Stephane Erard
Robin Fergason
Patricia Lawton
Xiaogang Ma

B.1. Leadership

NASA Review Manager: Rebecca McCauley Rench, Planetary Science Division, Science Mission Directorate, NASA Headquarters

NASA Review Deputy Manager: Meagan Thompson, Planetary Science Division, Science Mission Directorate, NASA Headquarters

Chair: Melissa McGrath, SETI Institute

CTS Task manager: Autumn Manecke

NRESS Support: Susan Borden, Latessa Tuck, Jason Silveira

B.2. Subcommittees

B.2.1. Archiving

Co-Chair: Moses Milazzo, Consultant, Other Orb LLC

Executive Secretary: Sarah Black, U.S. Geological Survey

NASA Ex-officio Members:

- Meagan Thompson, NASA AST, Engineer Program Management
- David Smith

Members:

- Kate Crombie, Indigo Information Services LLC
- Martha Maiden, Consultant
- Joseph Masiero, Caltech/IPAC
- Flora Paganelli, National Radio Astronomy Observatory Green Bank
- Corrine Rojas, Arizona State University
- Priyanka Sharma, Jet Propulsion Laboratory, California Institute of Technology
- Marshall Tabetah, Plane Space Design

B.2.2. Searching

Co-Chair: Emily Lakdawalla, The Lakdawalla Group, LLC

Executive Secretary: Katya Gilbo, Consultant

NASA Ex-officio Member: Lindsay Hays, NASA Program Officer & Deputy Program Scientist

Members:

- Julie Castillo-Rogez, Jet Propulsion Laboratory, California Institute of Technology
- Vandana Desai, Caltech/IPAC
- Mary Ann Esfandiari, Consultant, Cornell Technical Services
- Laszlo Kestay, U.S. Geological Survey Astrogeology Science Center
- Matthew Miller, Jacobs/NASA Johnson Space Center
- Lynn Neakrase, Planetary Data System Atmospheres Node, Department of Astronomy, New Mexico State University

B.2.3. Utilization

Co-Chair: Sebastien Besse, Aurora Technology for European Space Agency

Executive Secretary: Don Hood, Texas A&M University

NASA Ex-officio Member: Sarah Noble, Planetary Science Division, Science Mission Directorate, NASA Headquarters

Members:

- Alyssa Bailey, Arizona State University, MASTCAM-Z Mars 2020 Rover Team
- Ross A. Beyer, SETI Institute and NASA Ames Research Center

- Henry Hsieh, Planetary Science Institute
- David Mayer, U.S. Geological Survey Astrogeology Science Center
- Chase Million, Consultant, Million Concepts
- Danielle Wyrick, Southwest Research Institute, Space Science and Engineering Division

B.2.4. Mining & Automation

Co-Chair: Bruce Wilson, University of Tennessee and Oak Ridge National Laboratory

Executive Secretary: Hannah Dattilo, Fisk-Vanderbilt Master's to PhD Bridge Program

NASA Ex-officio Members:

- Megan Ansdell, Planetary Science Division, Science Mission Directorate, NASA Headquarters
- Kenneth Hansen, Planetary Science Division, Science Mission Directorate, NASA Headquarters

Members:

- Abigail Azari, Space Sciences Laboratory, University of California, Berkeley
- Benjamin Burnett, Architecture Technology Corporation
- Amitabha Ghosh, Consultant, Tharsis Inc. & Cornell Technical Services
- Jason Laura, U.S. Geological Survey
- Jennifer Shin, Consultant, 8 Path Solutions LLC
- Orenthal James Tucker, NASA, GSFC Research AST, Planetary Studies
- Alexandria Ware, LASP/University of Colorado

B.2.5. Inter-relational

Co-Chair: Caroline Coward, Jet Propulsion Laboratory, California Institute of Technology

Executive Secretary: Blake Mendoza, B&P Mendoza Enterprises, LLC

NASA Ex-officio Members:

- Steven Crawford, NASA AST, Technical Management
- Lucas Paganini, NASA Program Scientist

Members:

- Reta Beebe, Retired College Professor, New Mexico State University
- Shawn Brooks, Jet Propulsion Laboratory, California Institute of Technology
- Stephane Erard, PSL / Observatoire de Paris / LESIA
- Robin Fergason, U.S. Geological Survey Astrogeology Science Center
- Patricia Lawton, Planetary Data System Small Bodies Node, University of Maryland, College Park
- Xiaogang Ma, University of Idaho, Computer Science

C Subcommittees' Approaches

C.1. Archiving Subcommittee

The Archiving subcommittee focused on assessing the current state of archiving of NASA's planetary data. The data included in this assessment was not limited to science data acquired by missions, telescopes, etc. The assessment also included: "higher level" data that are produced by NASA PSD-funded scientists, researchers, and engineers; data that are produced by NASA-associated institutions; and data that are produced by citizen scientists, volunteers, and enthusiasts. To make the assessments, the Archiving subcommittee utilized public documents (for example, the PDS Roadmap from 2017); responses to the RFI (Appendix E); interviews with various data producers and repository managers; written responses to additional requests for information sent by the subcommittee or full committee; and the subcommittee members' own experiences and knowledge.

C.2. Searching Subcommittee

The Searching subcommittee assessed how well the Planetary Data Ecosystem serves its user communities by making its data holdings easy to discover, locate, navigate, browse, and retrieve. While the Planetary Data System is formally structured and has a full range of tools designed specifically for navigating and locating data within its nodes, the Ecosystem concept describes a larger, broader, more eclectic range of useful and relevant data sources that tend to serve smaller subcommunities and are not all easily discoverable (or even intelligible) by the full range of users.

The Searching subcommittee began with an attempt to understand the full landscape of the Planetary Data Ecosystem by assembling as complete a list as possible of its individual elements. The Searching subcommittee's work produced the definition of the Ecosystem in Chapter 2. Appendix I contains a more thorough -- though not exhaustive -- list of Ecosystem Elements.

In addition to mapping out the landscape of Ecosystem elements, the Searching subcommittee employed a concept of user personas to identify a broad range of potential users and characterize their experiences in attempting to answer specific research questions through data searches. Personas are hypothetical narratives intended to represent the context and potential search desires of real users. These narratives contain relevant, believable, specific, and precise details about their work role, goals, main tasks, usage stories, problems in work practice, concerns, barriers to their work, etc. Following are two example personas:

- **Persona Example 1:** Rose is a new mission operations staff member who needs to find the best base maps of the Moon on which to plan observations for a new high-spatial-resolution mapping spectrometer. Wavelength-dependent albedo, topography, and roughness are important because each affects the exposure and integration settings to obtain optimal data. She knows her instrument and the science of spectroscopy, but is not a lunar scientist.

- **Persona Example 2:** Sheetal is an illustrator who needs to create an infographic for a magazine article that shows all the asteroids in the solar system that have been visited by spacecraft or imaged by Earth-based radar. She also needs to include facts about their sizes and likely compositions. She is an experienced science illustrator, but does not have access to an academic library for reading paywalled publications.

The set of personas assembled by the Searching subcommittee represented as broad a spectrum of potential users as possible within the limitations of time and personnel available to the Board. Analysis of the personas list generated by members of the full Board yielded 6 major groups of personas who regularly search and generate data within the PDE. In general, the 6 groups listed below form a continuum in terms of their level of access to the process of generating mission data.

- **Mission team members** are data generators. They search the Ecosystem not only for data from their own missions, but also for data, instrument specifications, and results from other missions to the same or similar targets. Goals include planning future observations, creating derived products that follow established conventions, avoiding duplication of effort, and facilitating cross-correlation between the data sets they are producing and data sets produced by other missions.
- **Professional researchers** need to search the Ecosystem using a wide variety of sometimes esoteric parameters; to figure out whether their desired data set exists; and, if it does exist, to determine whether they have the best available version of a data set for their needs. They require tools that can filter and limit search results to a useful subset of large data sets. They need clear documentation that describes the data set and its limitations.
- **Graduate students** or similar early-career personnel are actively attempting to accumulate domain knowledge, but do not yet have expertise. Their searches may be inefficient without some form of guidance or assistance. They need easy connections to materials that will help them rapidly add to their expertise.
- **Media professionals** are working on a deadline, seeking information about missions and data in a way that is accessible to their audiences. Their searches usually crosscut many missions, and their preference is for synthesized and interpreted products rather than raw data sets. They do not have time to master individual data sets, and often do not need the best possible products; rather, they need to rapidly discover good enough materials for their purposes.
- **Professional educators** are creating teaching curriculum or training materials. They seek derived data sets that are easy to use and are described in plain language. They need data sets to remain in stable locations over the long term to ensure that educational materials do not become obsolete.
- **General space enthusiasts** do not always have specific search goals in mind -- they often wish to browse, rather than search. They need text in plain language, and pathways to appropriate resources to help them learn more.

Note that any one individual may fall into many of these persona groups in their different roles. For instance, mission team members are commonly also professional educators attempting to access other types of data, and are commonly space enthusiasts about fields outside their own.

The exercise provided a snapshot view into the successes and challenges encountered during searches as well as the number of steps required to locate data (or, in some cases, the number of steps a persona might take before giving up). In general, the exercise revealed that data sets and repositories were difficult to locate; even expert professionals could not readily find where goal data sets were supposed to be discoverable if they were not within the purview of the PDS. Every persona had to choose a place to begin a search, and most began with an Internet search engine like Google. Sites that host planetary data were generally not the top search results. Personae lacked confidence that the endpoints of their searches produced the most suitable data products. In many cases, a multi-step search for specific data did not result in the persona finding the optimum product for their research goal, an outcome that leads to duplicative effort and potentially lower-quality scientific output.

C.3. Utilization Subcommittee

The Utilization subcommittee focused on the tools and services that are used by experts and non-experts to evaluate, analyze and enjoy scientific observations returned by scientific experiments. A large fraction of the tools and services that the subcommittee evaluated relate to planetary mission products. The subcommittee spent significant effort considering the advantages and disadvantages of recommending an Open Source Software (OSS) policy for software developed under NASA Research and Analysis funding.

This initial work established context for the Utilization subcommittee to assess whether NASA Research and Analysis grant programs are relevant to, and supportive of, the diverse scientists and software developers that comprise the planetary science community. The subcommittee also evaluated NASA mechanisms that encourage the broader software developer community to build tools for planetary science, and developed recommendations for how to improve those mechanisms.

In the course of this assessment the Utilization subcommittee found it helpful to define three specific classes of utilization tools:

- **Planetary Data Ecosystem (PDE) tools.** Tools or utilization resources are considered to be in scope of the PDE if their nearly-exclusive use case is for the utilization of planetary data or if they contain a sub-component with a nearly-exclusive purpose for utilization of planetary data. For example, the JMARS software suite would be considered in scope of the PDE because its sole purpose (barring excessive creativity) is for the utilization of planetary data. For another example, while the ArcGIS and QGIS software systems are primarily used for terrestrial data, subcomponents of these tools have been modified to exclusively support planetary data, e.g. for non-terrestrial geoids. Consequently, ArcGIS and QGIS; so these tools are considered to be within the scope of the PDE. General-use software tools like text editors (e.g. Emacs), operating systems (e.g. Linux), or image processing suites (e.g. GIMP and Photoshop) are certainly critical to the use of planetary data, but they contain no functionality exclusive to the use of planetary data and are therefore out of scope.
- **Infrastructure tools.** Like roads, bridges, and power lines, infrastructure tools connect otherwise disparate elements of the PDE in a way that provides value to large cross-sections of planetary data users. Their cross-cutting nature also means that it is sometimes not clear that any single entity is primarily responsible (or “owns”) their development and maintenance, and it may therefore be necessary for NASA to directly initiate and support the development

and ongoing maintenance of these systems. Examples of extant infrastructure tools within the PDE include, but are not limited to, the following: Geospatial Information Systems for working with planetary mission data (perhaps specific capabilities added to QGIS); search capabilities across planetary mission data (including but not limited to the PDS); other software systems necessary to understand, process, calibrate, or use planetary data in general (e.g. SPICE, ISIS, GDAL, Ames Stereo Pipeline, Astropy, PlanetaryPy, etc.); and the PDS information model and associated software systems.

- **Community-produced tools.** Community-produced tools are those developed by stakeholders in the PDE in service to their own local interests or goals. Such tools may include software incidentally created in the course of a specific research project (which may or may not be supported by research funding), or by volunteer effort. Community-produced tools can have high impact and high value, but typically serve the needs of the stakeholders producing them and their close collaborators and are less responsive to the needs of others in the broader community.

C.4. Mining & Automation Subcommittee

The Mining and Automation (M&A) Subcommittee was charged with analyzing and assessing how emerging techniques, resources, and services support the PDE. The M&A Subcommittee focused on current and future best practices for a data ecosystem that should be brought to fruition for the PDE to stay relevant and effective at serving the community. This included ways that the PDE can better support the use of artificial intelligence, machine learning, advanced analytics methods, cloud computing, and automation to advance planetary science.

The committee started by reviewing all RFI responses (Appendix E) for relevant material, which directly led to three invitations for groups to meet with the subcommittee. The subcommittee also reviewed the material initially in the IRB document library, which also provided insight into areas for further investigation. Based on literature searches and committee member experience, several other documents were added to the IRB document library. Multiple subcommittee members attended sessions at the Fall 2020 AGU Virtual Meeting relevant to artificial intelligence, machine learning, advanced analytics, high end computing, automation, and cloud computing relevant to planetary science. Subcommittee members also attended a JPL-sponsored conference on machine learning and artificial intelligence, which had several sessions relevant to planetary science. Discussions with NASA personnel familiar with the committee's focus area led to invitations to two other groups to come speak with the subcommittee. The subcommittee also drew on their own personal experiences in the subject domain, as well as that of colleagues in their network.

This initial work led to a conclusion that there were three particular perspectives on these emerging techniques that the subcommittee needed to hear from:

> 1) Those with deep institutional connections to NASA planetary science, such as people at NASA centers and those in large, established research groups with a long history of NASA planetary science funding.

2) Those with more personal and individual connections to NASA planetary science, such as early career professionals, people working in small businesses (including individual consultants).

3) Those with limited or no current connections to NASA planetary science, but who have substantial expertise in these emerging techniques, resources, and services. This last group includes people who have applied these emerging techniques to planetary science, and have moved on to applications in other domains, often because of the technical and structural barriers which are discussed in this report.

The committee used a range of channels to solicit input from these groups, recognizing that the first two groups were relatively easy to reach. These channels included the outreach through the subcommittee members' professional networks, postings in on-line forums frequented by practitioners of these emerging techniques, an open on-line Town Hall meeting, and attendance at relevant on-line conferences.

Because Mining and Automation has a substantial degree of overlap with the domains of the other subcommittees, most Mining and Automation meetings included representatives of these other subcommittees. Mining and Automation subcommittee members also regularly attended meetings from the other subcommittees.

C.5. Inter-Relational Subcommittee

As content from the other subcommittees came forward, the Inter-relational subcommittee examined their work and revealed alliances, synergies, and complimentary comments across the IRB. The inter-relational subcommittee inventoried these common, overlapping items and synthesized them into higher level findings and recommendations that will make the proposed ecosystem more cohesive and clearly connected. These include identifying and defining ecosystem elements user types (or personas), areas missing from the discussion (such as cybersecurity, and data governance standards); evaluating training, education, and other external communications; examining non-U.S. and international data driven initiatives and agreements as exemplars; recommending pathways to evolve the PDE from its current state to an ideal state; and any other areas that boost the collaborative nature of the PDE.

C.6. Small Bodies Focus Group

As the IRB conducted its assessment of the Planetary Data Ecosystem, a few members of the IRB noted that, for historical and structural reasons, the PDE IRB discussions have been dominated by elements of the ecosystem associated with the solar system's largest bodies. An ad-hoc focus group self-organized with the objective of ensuring that the vast amount of data falling under the umbrella of "small bodies" was given sufficient attention by the IRB. The Small Bodies Focus Group assessed the state of archiving, searching, utilization, and automation of small-body data. The focus group contributed a non-exhaustive list of PDE elements associated with the small bodies to Appendix I. Focus group findings and recommendations arose mostly from the personal experience of various panel members.

D IRB Meeting and Presentation List

Date	Meeting Type	Speaker/Topic
11/12/2020	Full IRB	Kickoff meeting
11/16/2020	Utilization	Discussion of charge and ongoing schedule
11/17/2020	Searching	Personas
11/18/2020	Archiving	Subcommittee kickoff and organizational discussion
11/18/2020	Mining & Automation	
11/20/2020	Inter-Relational	
11/23/2020	Utilization	Plans for capturing findings as we meet, discussion of PDART program
12/02/2020	Searching	PDE Elements, RFIs, Personas, Planning month
12/03/2020	Steering	Subcommittee reports & breakout sessions, synthesis/upcoming plans
12/03/2020	Mining & Automation	
12/04/2020	Archiving	Speaker invitations, review PDS roadmap
12/07/2020	Utilization	Open Source Software and review of prior NAS report
12/07/2020	Searching	Steven Crawford: Open data at NASA
12/11/2020	Archiving	PDS roadmap discussion, questions for archival institutions
12/11/2020	Inter-Relational	
12/11/2020	Searching	Updates, Personas, Data Types, PDE Elements
12/14/2020	Utilization	Further discussion of Open Source Software
12/17/2020	Full IRB	Subcommittee reports & breakout sessions, synthesis/upcoming plans
12/18/2020	Archiving	Finish PDS roadmap discussion, finalize questions for PDS nodes
01/06/2020	Searching	Preliminary findings and recommendations
01/08/2021	Inter-Relational	Finalize RFI topic areas, begin crafting narratives to Appendix A questions
01/08/2021	Archiving	Discuss & draft speaker invitations, discuss missing datasets
01/11/2021	Utilization	Steven Crawford: NASA’s Current Open Software Policy and Future Changes
01/13/2020	Searching	
01/14/2021	Mining & Automation	
01/15/2021	Steering	Subcommittee updates, speaker invitations, final report, full IRB meeting plans
01/15/2021	Inter-Relational	
01/15/2021	Archiving	Review PDE elements draft, discuss final report structure
01/18/2021	Searching	Personas Discussion
01/21/2021	Full IRB	Personas, cross-cutting themes, utilization across the PDE, open science

01/22/2021	Archiving	Sebastien Besse: ESA Planetary Science Archive and Guest Storage Facility
01/25/2021	Utilization	Experience of Early Career Scientists and the PDE
01/26/2021	Steering	Final report outline
01/27/2021	Searching	
01/27/2021	Archiving	Finalize charter, draft outline structure
01/28/2021	Mining & Automation	
01/29/2021	Inter-Relational	
01/29/2021	Archiving	Kerstin Lehnert: AstroMat & Sample archiving
02/01/2021	Utilization	Closing discussion of PDART and OSS policy
02/02/2021	Archiving	Guest speaker preparation, updates from other meetings
02/03/2021	Searching	
02/04/2021	Mining & Automation	Chris Lynnes: Earthdata Cloud
02/04/2021	Inter-Relational	Bruce Wilson
02/05/2021	Archiving	Mike Nolan: Arecibo and other ground-based radar data preservation
02/08/2021	Mining & Automation	Jordan Padams: PDS Engineering Node Manager
02/08/2021	Utilization	Community training and pathways to fund infrastructural software tools
02/10/2021	Archiving	Prepare findings & recommendations for PDS briefing
02/10/2021	Searching	
02/11/2021	Mining & Automation	Frontier Development Laboratory
02/12/2021	Archiving	Emily Law: Treks platform
02/12/2021	Inter-Relational	Steve Joy & Baptiste Cecconi: IVOA and Solar System Interest Group / VESPA
02/17/2021	Steering	Subcommittee updates, full IRB meeting plans, PDS briefing plans
02/18/2021	Mining & Automation	Town Hall on AI/ML
02/19/2021	Archiving	Writing up findings & recommendations
02/19/2021	Mining & Automation	JPL Machine Learning Group
02/19/2021	Searching	
02/19/2021	Inter-Relational	Betsy Edwards, Cybersecurity
02/22/2021	Utilization	Formalization of findings thus far
02/23/2021	Full IRB (part 1)	Discuss PDS-related findings & recommendations
02/24/2021	Full IRB (part 2)	Discuss other findings & recommendations
02/26/2021	Archiving	Full IRB meeting debrief, review findings & recommendations
02/26/2021	Steering	Assembling PDS briefing presentation
03/01/2021	Utilization	Discussion of final findings & recommendations
03/02/2021	PDS briefing	Present initial PDS-related findings and recommendations to NASA/PDS

03/02/2021	Archiving	Discuss PDS briefing
03/04/2021	Utilization	Sarah Noble: NASA’s PDART program and its trajectory
03/04/2021	Mining & Automation	
03/05/2021	Steering	Discuss plans for final report writing
03/08/2021	Utilization	Final meeting and wrap-up
03/09/2021	Writing Group	Writing group check in
03/10/2021	Archiving	Check in: writing up supporting information for findings & recommendations
03/11/2021	Searching	NASA Acting Chief Archivist Holly McIntiyre on NASA Center Archives
03/11/2021	Mining & Automation	
03/12/2021	Writing Group	Writing group check in
03/16/2021	Writing Group	Writing group check in
03/16/2021	Archiving	Updates from writing committee, upcoming tasks, wrap up
03/18/2021	Searching	
03/19/2021	Writing Group	Writing group check in
03/23/2021	Writing Group	Writing group check in
03/26/2021	Writing Group	Writing group check in

E Disposition of Responses to Request for Information

No.	Title	Disposition
1	Need for Infrastructure of Planetary Data Training	This Response identifies a need for tools training addressed by Finding 61, Finding 67, Recommendation 50, Recommendation 64, and Recommendation 65. Response identifies needs for higher-level science-ready data products, addressed by Finding 60 and Recommendation 49. Response further advocates for funding a particular program, but this report does not address funding recommendations at that level of granularity.
2	Need for Infrastructure of Unique Planetary Data Digitization	This Response, like many others, pointed out that there are several data sets that do not have a home in the Ecosystem. The general issue of many data sets that do not have a clear home in the PDE is discussed in many parts of the report but most directly in Finding 45 and Recommendation 29. However, this Response is focused on the fate of the data from the former Regional Planetary Image Facilities. These data are prime examples of materials at risk of irretrievable loss, covered by all of Section 4 but most specifically by Finding 47, Finding 52, Finding 53, Recommendation 28, Recommendation 29, and Recommendation 30.
3	The Need for a Planetary Modeling Output Archive Infrastructure	This Response suggests that NASA work to ensure there is an archive that can accommodate modeling output. It is addressed by Finding 55 and Recommendation 44.
4	Creating an ozone layer on Mars using natural processes	Not responsive to the RFI.
5	Planetary Data System Cloud Status and Plans	This Response is a summary of the PDS' current plans for the use of cloud technologies and is addressed in Section 3.2.8.
6	The Planetary Data System Data Services Initiative	This Response is a broad review of the PDS, addressed throughout this report, but especially in Section 3.
7	Need for Easy-To-Use Planetary Orbital Mapping Tools	This Response identifies the need for higher-level science-ready data products, addressed by Finding 60 and Recommendation 49. Response identifies the value of open-source software, addressed by Finding 66, Recommendation 62, and Recommendation 63. Response identifies needs for foundational data products addressed by Finding 30, Recommendation 19, and Recommendation 20. Response further advocates for specific activities, but this report does not address funding recommendations at that level of granularity.

8	A Strategy for Managing NASAs Long Tail of Planetary Research Data: Insights from the Development of the Astrobiology and Habitable Environments Database (AHED).	This Response summarizes the recommendations from Strategic Data Management Working Group final report through the experience of building the Astrobiology and Habitable Environments Database. It is addressed in many locations, but its three specific recommendations are most directly addressed by Finding 10, Finding 11, Finding 13, Finding 46, Recommendation 7, Recommendation 8, Recommendation 10, and Recommendation 28.
9	Response to RFI: Planetary Data Ecosystem (NNH20ZDA011L)	This Response advocates for better informing grant proposers of available software products and applicable costs (if any) that could be used for their investigations, availability of free trial periods for testing out commercial PDE-related software products, and a means for community members to provide computer-based science/engineering solutions to address PDE needs. Recommendation 42 addresses the need for a comprehensive software archiving strategy, and Section 6.1.2. addresses the other software-related concerns raised in the Response.
10	Comment on Impediments to Effective and Efficient PDE Data Usage	This Response advocates for NASA to create a Venus data analysis program and points out specific issues with access to foreign mission data and legacy data sets. We address these in Finding 6 and Recommendation 4. We do not address issues that were raised with specific R&A programs in this report.
11	Comment on Impediments to Maintenance of High-functioning Computational Tools for Groundbreaking Planetary Science Work	This Response describes a number of issues related to the support and maintenance of software tools in the specific context of Venus atmosphere models. The broad software issues are addressed in our recommendations related to the funding and preservation of software. Some findings and recommendations regarding this include Finding 65, Recommendation 58, Recommendation 59, and Recommendation 60.
12	Strategies for a Forward-Looking Planetary Data Ecosystem	This Response provides general strategies in the context of Solar System Trek as an element of the PDE. The overall strategies discussed in the response are aligned with the report. Some specific issues raised in the response are addressed in Finding 51, Finding 52, Finding 54, Recommendation 37, Recommendation 39, and Recommendation 42.
13	Interdisciplinary Extrasolar Planetary Data Access and Research	This Response details how exoplanets fall in a gap at NASA between astrophysics and planetary science. It is addressed in Finding 6, Finding 11, Finding 59, Recommendation 4, Recommendation 8, Recommendation 48, Recommendation 49, Recommendation 50, and Recommendation 52.

14	Planetary Analog Field Data: Present and Future Accessibility and Discoverability	This Response made three specific recommendations with relation to terrestrial analog data. These are addressed in Finding 10, Finding 11, Finding 50, Recommendation 7, Recommendation 8, and Recommendation 36.
15	Establishing Mission Context to Enhance Team Science	This Response highlighted the need to preserve the context for mission data with an emphasis on metadata standards that would facilitate a consistent modern API to facilitate data access. It is addressed most specifically in Finding 13, Finding 23, Finding 41, Recommendation 10, Recommendation 14, and Recommendation 26.
16	Foundational Software Components in Planetary Science	This Response identifies categories of “foundational software components” and advocates for ongoing support of open-source software for planetary science. Addressed in Finding 54, Finding 60, Recommendation 42, Recommendation 43, Recommendation 49, and Section 6.1.2.
17	The importance of foundational spatial data products to a healthy Planetary Data Ecosystem	This Response highlights the importance of “foundational” spatial data products especially in allowing co-analysis of different data sets. It is addressed most specifically in Finding 30, Recommendation 20, and Recommendation 21.
18	The Role of Metadata in a Planetary Data Ecosystem	This Response highlights the essential role that metadata plays in allowing a de-centralized community to operate effectively. It is addressed most directly in Finding 13 and Recommendation 10.
19	The importance of coordinate system standards to a healthy Planetary Data Ecosystem	This Response highlights the importance of coordinate system and other standards being considered early in mission planning. It is addressed most directly in Finding 30 and Recommendation 21.
20	Data Available to Planetary Community	This Response is generally addressed throughout Section 3 and Section 4, and most directly in Finding 29, Finding 40, Finding 41, Recommendation 19, and Recommendation 26.
21	Frontier Development Lab Planetary Data Ecosystem Response	This Response led to a speaker invitation on 11 Feb 2021 and the Artificial Intelligence/Machine Learning Town Hall on 18 Feb 2021. Comments received from the community were incorporated into Section 6.
22	Mapping FAIR Themes Across a Planetary Data Ecosystem	This Response provides a summary of FAIR data principles and how they could be applied to planetary data. FAIR data principles were central to many of the IRB discussions. This response was addressed throughout the report but most directly in Finding 14, Recommendation 11, Finding 53, Recommendation 16, Recommendation 29, and Recommendation 41.
23	The Need for Improved Integration of Small Solar System Body Data Sources	This Response highlighted a specific example of how the concept of the PDE provides a framework for important

		new science. This example was included in the report as a case study in Section 7.2.2.
24	The Outer Planets Unified Search (OPUS) and its place within the Planetary Data Ecosystem	This Response highlights one element of the PDE, the Outer Planets Unified Search tool. This element was included in the list of PDE elements (Appendix I). It is most directly addressed in Finding 17 and Recommendation 12.
25	The PDS Ring-Moon Systems Node and its place within the Planetary Data Ecosystem	Duplicate with Response 24.
26	Magnetic Data Archives for Returned Astromaterials and Meteorites	This Response highlights a gap in the current PDE when it comes to magnetic data collected from astromaterials. It is addressed most directly in Finding 49, Recommendation 33, and Recommendation 34.
27	The NASA Planetary Data Systems Small Bodies Nodes New Challenges and Cloud Solutions	This Response highlighted a specific use case for cloud computing for analysis of data related to small bodies. Led to example presented in Section 7.2.2. Portions are addressed in Section 3.2.8.
28	Eliminating redundancies enables high value activities that enable ecosystem strategic goals	This Response was very broad in scope and is relevant to many of the IRB’s activities. As such, it is addressed throughout the report. Some of the most direct examples include Finding 28, Finding 30, Finding 54, Recommendation 17, Recommendation 18, Recommendation 21, Recommendation 42, and Recommendation 43.

F Reference Material

ADA (1982) Anti-Deficiency Act, Public Law 97-258, https://www.govinfo.gov/content/pkg/STATUTE-96/pdf/STATUTE-96-Pg877.pdf

Archinal, B. and the Lunar Geodesy and Cartography Working Group (2009) Activities of the NASA LPRP Lunar Geodesy and Cartography Working Group, *Lunar and Planetary Science Conference* 40, #2095. See also https://www.usgs.gov/centers/astrogeology-science-center/science/nasa-planetary-cartography-planning?qt-sci

Archinal, B. A., Acton, C. H., A'Hearn, M. F., Conrad, A., Consolmagno, G. J., et al. (2018) IAU Working Group on Cartographic Coordinates and Rotational Elements: 2015, *Celest Mech Dyn Astr*, 130:22 https://doi.org/10.1007/s10569-017-9805-5

Astrobiology Habitable Environments Database (AHED) https://www.nasa.gov/ames/ahed

Azari, A. R., Biersteker, J. B., Dewey, R. M., Doran, G., Forsberg, E. J., et al. (2020) Integrating Machine Learning for Planetary Science: Perspectives for the Next Decade, *White Paper to the Planetary Science and Astrobiology Decadal Survey* 2023-2032, https://arxiv.org/abs/2007.15129

Beyer, R., Hare, T., Radebaugh, J. (2018) The Need for a Planetary Spatial Data Clearinghouse, *Informatics and DataAnalytics Workshop*, #6067, https://wufs.wustl.edu/psida/2018/abspres/pdf/6067.pdf

Borowitz, M. (2019) Government Data, Commercial Cloud: Will Public Access Suffer? *Science* 363, 588-589, doi: 10.1126/science.aat5474. https://science.sciencemag.org/content/363/6427/588

Bristow, T.F. et al. (2020) Strategy for Managing NASA's Long Tail of Planetary Research Data Insights from the Development of the Astrobiology and Habitable Environments Database (AHED). *White paper to the Planetary Science and Astrobiology Decadal Survey* 2023-2032. http://surveygizmoresponseuploads.s3.amazonaws.com/fileuploads/623127/5489366/58-f01767cdd5ccb5055262989bf3b320e5_BristowThomasF.pdf

Office of Management and Budget (1993) *Circular A-110*, Revised 11/19/93, As Further Amended 9/30/99 https://obamawhitehouse.archives.gov/omb/circulars_a110/

Collection of PSDI publications (1981-2014) https://astrogeology.usgs.gov/groups/nasa-planetary-cartography-planning, ftp://ftpext.usgs.gov/pub/wr/az/flagstaff/barchinal/Carto-Planning/

Cooperative Agreement Notice (CAN) NNH15ZDA006C (released 5 March 2015 and due 1 June 2015). https://nspires.nasaprs.com/external/viewrepositorydocument/cmdocumentid=449464/solicitationId=%7B6C202F60-D0E1-91E6-0D59-A6FC8C96EA36%7D/viewSolicitationDocument=1/PDSCAN Amend1.pdf

Coretrustseal Repository Certification Requirements (2020) Coretrustseal.Org

Duxbury, T. C., Kirk, R. L., Archinal, B. A., Neumann, G. A. (2002) Mars Geodesy/Cartography Working Group Recommendations on Mars Cartographic Constants And Coordinate Systems, Symposium on Geospatial Theory, Processing and Applications, *Symposium sur la théorie, les traitements et les applications des données Géospatiales*, https://www.isprs.org/proceedings/XXXIV/part4/pdfpapers/521.pdf

ESDSWG, Data Product Development Guide for Data Producers, https://earthdata.nasa.gov/esdis/eso/standards-and-references/data-product-development-guide-for-data-producers

Executive Office of the President Office of Management and Budget (2013) *Memorandum for the Heads of Executive Departments and Agencies, Subject: Open Data Policy-Managing Information as an Asset*. Https://Obamawhitehouse.Archives.Gov/Sites/Default/Files/Omb/Memoranda/2013/M-13-13.Pdf

FAIR, https://www.go-fair.org/fair-principles/; https://en.wikipedia.org/wiki/FAIR_data

Friesenhahn, I. and Beaudry, C. (2014) *The Global State of Young Scientists – Project Report and Recommendations*, Berlin: Akademie Verlag, ISBN: 978-3-939818-44-1

Hartson, R., Pyla, P. S. (2012) *The UX book: Process and guidelines for ensuring a quality user experience*. Amsterdam, The Netherlands: Morgan Kaufmann.976 pp.

Hasselbring, W., Carr, L., Hettrick, S., Packer, H., Tiropanis, T. (2020) From FAIR research data toward FAIR and open research software, *it - Information Technology* 62, 39-47, https://doi.org/10.1515/itit-2019-0040

OSTP Director Holdren, J. P. (2013) Memorandum for the Heads of Executive Departments and Agencies, Subject: Increasing Access to the Results of Federally Funded Scientific Research 1. Policy Principles, February 22, 2013, https://Obamawhitehouse.Archives.Gov/Sites/Default/Files/Microsites/Ostp/Ostp_Public_Access_Memo_2013.Pdf

IAU Working Group Website (2020) https://www.iau.org/science/scientific_bodies/working_groups/

ISO 9241-210, https://en.wikipedia.org/wiki/ISO_9241#ISO_9241-210

Johnson, J. R., Archinal, B., Kirk, R., Gaddis, L., Anderson, J., and the NASA Planetary Cartography and Geologic Mapping Working Group (PCGMWG) (2010) The Importance of A Planetary Cartography Program: Status and Recommendations for NASA 2013–2023, *White paper to the Planetary Science Decadal Survey* 2013-2022.

JPL-D76488 (2015) TRK-2-34, DSN Tracking System Data Archival Format, DSN No. 820-013,TRK-2-34, Rev. O, https://pds-ppi.igpp.ucla.edu/data/maven-rose-raw/document/maven_rose_tnf_v01_r00.pdf

Laura, J.R., Hare, T.M., Gaddis, L.R., Fergason, R.L., Skinner, J.A., et al. (2017) Towards a Planetary Spatial Data Infrastructure, *ISPRS Int. J. Geo-Inf. 6*, 181. https://doi.org/10.3390/ijgi6060181

Laura, J., Bland, M.T., Fergason, R.L., Hare, T.M., Archinal, B.A. (2018) Framework for the Development of Planetary Spatial Data Infrastructures: A Europa Case Study, *Earth and Space Science* 5, 486-502, https://doi.org/10.1029/2018EA000411

Laura, J. R., Mapel, J., Hare, T. (2020) Planetary Sensor Models Interoperability Using The Community Sensor Model Specification, *Earth and Space Science* 7, e2019EA000713, https://doi.org/10.1029/2019EA000713

Lin, D., Crabtree, J., Dillo, I., et al. (2020) The TRUST Principles for digital repositories, *Sci Data* 7**,** 144. https://doi.org/10.1038/s41597-020-0486-7

Maskey, M., H. Alemohammad, K. J. Murphy, and R. Ramachandran (2020) Advancing AI for Earth science: A data systems perspective, *Eos, 101,* https://doi.org/10.1029/2020EO151245.

Nardi, B., and ‘Day, V (1999) *Information Ecologies: Using Technology with Heart*, https://mitpress.mit.edu/books/information-ecologies.

NASA, Frequently Asked Questions (FAQ) Page for ROSES About Data Management Plans (DMPs) (2021) https://Science.Nasa.Gov/Researchers/Sara/Faqs/Dmp-Faq-Roses

NASA (2014) NASA Plan for Increasing Access to the Results of Scientific Research, *Digital Scientific Data and Peer-Reviewed Publications*, December 2014, 22 pp. https://Science.Nasa.Gov/Science-Red/S3fs-Public/Atoms/Files/Nasa_Plan_For_Increasing_Access_To_Results_Of_Federally_Funded_Research1.Pdf

NASA SMD's Strategic Data Management Working Group (2019) *Science Mission Directorate's Strategy for Data Management and Computing for Groundbreaking Science 2019-2024*, https://science.nasa.gov/science-red/s3fs-public/atoms/files/SDMWG%20Strategy_Final.pdf

NASEM (2018) National Academies of Sciences, Engineering, and Medicine (2018) *Open Source Software Policy Options for NASA Earth and Space Sciences*, Washington, DC: The National Academies Press. https://doi.org/10.17226/25217

NASA Office of Inspector General (2020) *NASA's Management of Distributed Active Archive Centers*, Report No. IG-20-011, https://oig.nasa.gov/docs/IG-20-011.pdf

National Archives, What’s an National Archives? https://www.archives.gov/about/info/whats-an-archives.html

National Research Council (1982) *Data Management and Computation: Volume 1: Issues and Recommendations*. Washington, DC: The National Academies Press. https://doi.org/10.17226/19537.

National Geological and Geophysical Data Preservation Program, https://www.usgs.gov/core-science-systems/national-geological-and-geophysical-data-preservation-program/physical-sample

National Institute of Heath, NIH Clinical Research Trials and you, https://www.nih.gov/health-information/nih-clinical-research-trials-you/list-registries

NIST (2017) Digital Identity Guidelines, https://pages.nist.gov/800-63-3/

NPD 2200.1, Management of NASA Scientific and Technical Information, https://nodis3.gsfc.nasa.gov/displayDir.cfm?t=NPD&c=2200&s=1D

NPR 7120.e, NASA Space Flight Program and Project Management Requirements, https://nodis3.gsfc.nasa.gov/displayDir.cfm?t=NPR&c=7120&s=5E

Paganelli, F., Archinal, B. A., Acton, C. H., Conrad, A., Duxbury, T. C., et al. (2020) The Need for Recommendations in Support of Planetary Bodies Cartographic Coordinates and Rotational Elements Standards, *White paper to the Planetary Science and Astrobiology Decadal Survey* 2023-2032, http://surveygizmoresponseuploads.s3.amazonaws.com/fileuploads/623127/5489366/111-b0b92cc5c22150f768c0fca03cf31b70_FloraPaganelli.pdf.

PDART, Planetary Data Archiving, Restoration, and Tools program, https://nspires.nasaprs.com/external/solicitations/summary.do?solId=%7b451ABBBE-E682-9AF0-393F-8F30D0E0AD3B%7d&path=&method=init

PDS4 Data Provider’s Handbook: https://pds.nasa.gov/datastandards/documents/dph/

PDS Project Office (2020) PDS Customer Satisfaction Survey 2020 Presentation, https://pds.nasa.gov/home/NASAPDSUserSatisfaction2020final2.pdf

PDS Dictionary: https://pds.nasa.gov/datastandards/dictionaries/index-1.15.0.0.shtml#create-dictionary

PDS Roadmap Study (2017) NASA's *Planetary Data System Roadmap Study for 2017 – 2026*, https://pds.nasa.gov/home/about/PlanetaryDataSystemRMS17-26_20jun17.pdf

Prockter, L., Tiscareno, M., Grayzeck, E. J., Acton, C. H., Arvidson, R. E., et al. (2020) The Planetary Data System: A Vital Component In Nasa's Science Exploration Program, *White Paper to the Planetary Science and Astrobiology Decadal Survey 2023-2032* https://www.seti.org/sites/default/files/2020-09/PDS%20WP%20FINAL.pdf

Radebaugh, J., Thomson, B. J., Archinal, B., Beyer, R., DellaGiustina, D., et al. (2020) Maximizing the value of Solar System data through Planetary Spatial Data Infrastructures, *White Paper to the Planetary Science and Astrobiology Decadal Survey 2023-2032,* https://arxiv.org/abs/2008.06171

Ramapriyan, H. K. and Behnke, J. (2020) NASA's Earth Observing System Data and Information System (EOSDIS) and FAIR – A Self-Assessment. *American Geophysical Union 2020 Fall Meeting*, #IN044-08. https://agu.confex.com/agu/fm20/meetingapp.cgi/Paper/683318.

Ramapriyan, H. K., Leonard, P. J. T. (2020) Data Product Development Guide (DPDG) for Data Producers version1, *NASA Earth Science Data and Information System Standards Office*, 9 July 2020. https://doi.org/10.5067/DOC/ESO/RFC-041VERSION1

ROSES (2021) https://science.nasa.gov/researchers/sara/grant-solicitations, Guidebook for Proposers: https://www.nasa.gov/sites/default/files/atoms/files/2021_ed._nasa_guidebook_for_proposers.pdf

Skinner, J.A. Jr., Huff, A.E., Fortezzo, C.M., Gaither, T., Hare, T.M., Hunter, M.A., Buban, H. (2019) *Planetary geologic mapping—program status and future needs: U.S. Geological Survey Open-File Report* 2019–1012, 40 p., https://doi.org/10.3133/ofr20191012

Society of American Archivists, What are Archives? https://www2.archivists.org/about-archives

University Consortium of Geographic Information Science (2006) *Geographic Information Science & Technology Body of Knowledge*, Eds. DiBiase, D., Demers, M, Johnson, A., Kemp, Assn of Amer Geographers; 1st edition (June 30, 2006), 162 pp., https://www.ucgis.org/gis-t-body-of-knowledge

White House, Office of Management and Budget (2021) Electronic Code of Federal Regulations Part 200: Uniform Administrative Requirements, Cost Principles, and Audit Requirements for Federal Awards. https://www.ecfr.gov/cgi-bin/text-idx?tpl=/ecfrbrowse/Title02/2cfr200_main_02.tpl

White House, Office of the Press Secretary (2013) Executive Order -- Making Open and Machine Readable the New Default for Government Information, released May 09, 2013, https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-

Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3**,** 160018. https://doi.org/10.1038/sdata.2016.18

Williams, D. A., Anderson, R. C., Byrne, S., Costard, F., Hayes, A., Jaumann, R., et al. (2020) RPIFs to PDUCs: New Planetary Data Utilization Centers to support NASA's Planetary Data Ecosystem, *White Paper to the Planetary Science and Astrobiology Decadal Survey 2023-2032*, https://repository.hou.usra.edu/bitstream/handle/20.500.11753/1679/LPI-002441.pdf?sequence=1&isAllowed=y

Williams, D. A. (2021) Request for Information to Archiving Committee, Planetary Data Ecosystem Independent Review Board.

G Acronym List

3D: three-dimensional

AA: Advanced Analytics

ACSI: American Customer Satisfaction Index

ACT: Applied Coherent Technology Corporation

ADS: Astrophysics Data System

AGU: American Geophysical Union

AHED: Astrobiology Habitable Environments Database hosted by NASA ARC

AI: Artificial Intelligence

AMMOS: Advanced Multi-Mission Operations System

AO: announcement of opportunity

API: application programming interface

APL: Applied Physics Laboratory, part of John Hopkins University

APPS: AMMOS-PDS Pipeline Service

ARC: NASA Ames Research center

ARD: analysis-ready data

ARES: NASA Johnson Space Center’s Astromaterials and Exploration Science Division

ARMS: AHED’s Astrobiology Resource Metadata Standard

ASCII: American Standard Code for Information Interchange

AWS: Amazon Web Services

CADC: Canadian Astronomy Data Centre

CAPTEM: Curation and Analysis Planning Team for Extraterrestrial Materials, recently renamed Extraterrestrial Materials Analysis Group

CASRAI: Consortia Advancing Standards in Research Administration Information

CATCH: Comet Asteroid Telescopic Catalog Hub

CCSDS: Consultative Committee for Space Data Systems

CDAWeb: Coordinated Data Analysis (Workshop) Web

CDF: Common Data Format

CD-ROM: Compact Disc Read-Only Memory

CEOS: Committee on Earth Observing Satellites

CEPS: The Smithsonian National Air and Space Museum’s Center for Earth and Planetary Studies data repository

CfA: Center for Astrophysics | Harvard & Smithsonian

CIS-JPL: Cartography and Image Science Node of the PDS at JPL

CIS-USGS: Cartography and Image Science Node of the PDS at the USGS

CNEOS: Center for Near Earth Object Studies at JPL

CNRS: Centre National de la Recherche Scientifique

CODMAC: National Research Council Committee on Data Management and Computation

CRISM: Compact Reconnaissance Imaging Spectrometer for Mars

CRAG: NASA’s Cartography Research Assessment Group

CSV: comma-separated values

CTS: Cornell Technical Services

DAAC: Distributed Active Archive Center, part of EOSDIS

DAP: Data Analysis Program

DB: database

DBaaS: Database-as-a-Service (ARC’s Nebula Cloud Computing Platform)

DMP: Data Management Plan

DN: Digital Number

DOI: digital object identifier

DPDG: data product development guide

DPS: Division for Planetary Sciences, part of the American Astronomical Society

DSN: Deep Space Network

DVD: Digital Video (or Versatile) Disc

ECCO: NASA’s Estimating the Circulation and Climate of the Ocean consortium

EN: Engineering Node of the PDS

EOSDIS: Earth Observing System Data and Information System

ESDIS: Earth Science Data and Information System project.

ESA: European Space Agency

ESD: NASA Earth Science Division

ESDIS: Earth Science Data and Information System (the project that runs EOSDIS)

ESDS: NASA Earth Science Data Systems (the program which funds ESDIS)

ESDSWG: Earth Science Data Systems Working Group

ESIP: Earth Science Information Partners

ESO: European Southern Observatory

FAIR: findability, accessibility, interoperability, and reusability

FAQ: frequently asked questions

FGDC: Federal Geographic Data Committee

FITS: Flexible Image Transport System

GDAL: Geospatial Data Abstraction Library, gdal.org

GIBS: Global Imagery Browse Services

GIMP: GNU Image Manipulation Program

GIS: geospatial information system

GNU: recursive acronym for “GNU's Not Unix!”

GSFC: NASA/Goddard Space Flight Center

HAPI: Heliophysics API

HDF: Hierarchical Data Format

HEC: high-end computing

HECC: NASA’s High-End Computing Capability

HELIO: NASA’s Heliophysics Division

HEOMD: Human Exploration and Operations Mission Directorate

HiRISE: High Resolution Imaging Science Experiment

HITRAN/Harvard: High-Resolution Transmission Molecular Absorption Database, hosted by the Harvard-Smithsonian Center for Astrophysics

HRP: NASA’s Human Research Program

HQ: Headquarters

HSD: NASA/GSFC’s Heliophysics Science Division

IAA: International Academy of Astronautics

IaaS: Infrastructure-as-a-Service (ARC’s Nebula Cloud Computing Platform)

IAU: International Astronomical Union

IAWN: International Asteroid Warning Network

ID: identity document

IDP: Interplanetary Dust Particles

IHDEA: International Heliophysics Data Environment Alliance

IP: Internet protocol

IPAC: Infrared Processing and Analysis Center

IPDA: International Planetary Data Alliance

IRB: Independent Review Board

IRSA: NASA IPAC Infrared Science Archive

ISIS: Integrated Software for Imagers and Spectrometers

ISO: International Organization for Standardization

ISRO: Indian Space Research Organization

IVOA: International Virtual Observatory Alliance

JAXA: Japan Aerospace Exploration Agency

JMARS: Java Mission-planning and Analysis for Remote Sensing

JPEG: Joint Photographic Experts Group that created and maintains the JPEG standards

JPL: Jet Propulsion Laboratory

JSC: Johnson Space Center

JSON: JavaScript Object Notation

K-12: from kindergarten to 12th grade,

LDDTool: Local Data Dictionary Tool

LDSA: Life Sciences Data Archive

LPI: Lunar Planetary Institute

LPRP: NASA’s Lunar Precursor and Robotic Program

LROC: Lunar Reconnaissance Orbiter Camera

LSST: Large Synoptic Survey Telescope, now called Vera C. Rubin Observatory

M3: The Moon Mineralogy Mapper, an instrument on Chandrayaan-1

MAPSIT: Mapping and Planetary Spatial Infrastructure Team

MAST: Mikulski Archive for Space Telescopes

ML: machine learning

MPC: Minor Planet Center at CfA, a subnode of PDS SBN

MRCTR: The USGS Astrogeology Mapping, Remote-sensing, Cartography, Technology, and Research GIS Lab

NAIF: Navigation and Ancillary Information Facility Node of the PDS at JPL

NAS: NASA’s Advanced Supercomputing Division

NASA: National Aeronautics and Space Administration

NAT: Network Address Translation

NCCS: NASA Center for Climate Simulation

NED: National Elevation Dataset

NEOWISE: Near Earth Object Wide-field Infrared Survey Explorer

NetCDF: Networked Common Data Format

NIH: the United States National Institutes of Health

NRESS: NASA Research and Educational Support Services

NSF: National Science Foundation

NSPIRES: NASA Solicitation and Proposal Integrated Review and Evaluation System

NSSDCA: NASA Space Science Data Coordinated Archive

OCIO: NASA’s Office of the Chief Information Officer

ODE: Orbital Data Explorer hosted at the PDS Geosciences Node

OLAF: On-Line Archiving Facility

OMB: the United States Office of Management and Budget

OPUS: Outer Planets Unified Search

ORR: Operational Readiness Review

OSIRIS-REx: Origins, Spectral Interpretation, Resource Identification, Security, Regolith Explorer
OSTP: The White House Office of Science and Technology Policy
PaaS: Platform-as-a-Service (ARC’s Nebula Cloud Computing Platform)
PAC: Planetary Science Advisory Committee
Pan-STARRS: Panoramic Survey Telescope and Rapid Response System
PCGMWG: NASA Planetary Cartography and Geologic Mapping Working Group (1994-2012)
PDART: Planetary Data Archiving and Tools ROSES program
PDCO: NASA’s Planetary Defense Coordination Office
PDE: Planetary Data Ecosystem
PDF: Portable Document Format
PDS: Planetary Data System
PDS3: Version 3 of the PDS standard
PDS4: Version 4 of the PDS standard
PEN: Planetary Exploration Newsletter
PI: principal investigator
PILOT: Planetary Imaging LOcator Tool (PILOT)
PLAID: PDS Label Assistant for Interactive Design
PLRA: Program-Level Requirements Appendix
PMC: PubMed Central system of the NIH
PNG: portable network graphics image file format
PPI: Planetary Plasma Interactions Node of the PDS
PSA: ESA Planetary Science Archive
PSD: NASA Planetary Science Division
PSDI: planetary spatial data infrastructure, modeled after SDIs in use for Earth
PSIDA: Planetary Science Informatics and Data Analytics
QuAIL: NASA Quantum Artificial Intelligence Laboratory
R&A: research and analysis
RBSP-ECT: Radiation Belt Storm Probes mission Energetic particle, Composition, and Thermal plasma Suite
RDA: Research Data Alliance
RELAB/Brown: Reflectance Experiment Laboratory, hosted by Brown University
RFI: Request for information
RMS: Ring-Moon Systems (PDS Node)
ROSES: Research Opportunities in Space and Earth Science
RPIF: Regional Planetary Image Facility
SaaS: Software-as-a-Service (ARC’s Nebula Cloud Computing Platform)

SAO: Smithsonian Astrophysical Observatory

SBIR/STTR: Small Business Innovation Research and Small Business Technology Transfer programs

SBN: Small Bodies Node of the PDS

SDAC: NASA/GSFC Solar Data Analysis Center

SIMBA: System for Image Management extended By Analysis, an image management and analysis system hosted by Cornell’s Vision and Imaging Analysis Lab

SMBT: Small Body Mapping Tool

SMD: NASA Science Mission Directorate

SOC: science operations center

SPDF: NASA’s Space Physics Data Facility

SPICE: Spacecraft ephemeris, Planet, satellite, comet, or asteroid ephemerides, Instrument information, Orientation information (C-kernel), Events information; information system part of NASA’s Navigation and Ancillary Information Facility (NAIF)

SQL: Structured Query Language

SSCWeb: Satellite Situation center System and Services

SSEDSO: Solar System Exploration Data Services Office at NASA’s GSFC

SSIG: Solar System Interest Group of the IVOA

SSOIS: Solar System Object Image Search, a searching tool offered by the CADC

STEM: science, technology, engineering, and mathematics

STI: NASA’s Science and Technical Information Program

TESS: NASA’s Transiting Exoplanet Survey Satellite

THEMIS: Thermal Emission Imaging System

TIFF: Tagged Image File Format

TIR/ASU: thermal infrared spectral library hosted by Arizona State University

TRAPPIST: The Transiting Planets and Planetesimals Small Telescope

TSV: Titan Swatch Viewer

UMD: University of Maryland

UPC: Universal product code

URL: Uniform Resource Locator

USGS: United States Geological Survey

USRA: Universities Space Research Association

ViRBO: Virtual Radiation Belt Observatory

VMO: Virtual Magnetospheric Observatory

VIS-NIR/USGS: Visible and Near Infrared Spectral Library at USGS

VO: Virtual Observatory (falls under IVOA)

WDS: World Data System

WMS: Web Map Service

XML: Extensible Markup Language

ZTF: Zwicky Transient Facility

H Glossary

analysis-ready data: Important in the context of usability, mining, and automation, “ana;ysis-ready” describes data that is 1) accessible to users without much background knowledge or calibration requirements and 2) described with sufficient metadata/user guides to reproduce or undertake any processing needed for higher level analyses.

archive (digital): Digital archives have the following properties: they are independent, or managed by someone other than a major data provider; sustainable, managed for the long term (25 years at least); searchable; citable; preeminent, or considered by its user community as the “standard” archive for the subfield; and standardized, providing data products in standardized formats and file types. In addition, it is desirable for archives to be open-access and to provide peer review and documentation (user guides, calibration descriptions, etc.) for the data they hold. This definition comes from language in NASA’s Planetary Data Archiving, Restoration and Tools (PDART) announcements of opportunity. (Copied from Section 2.1.3.2)

archive (physical): A physical archive is a collection of historically significant items that are selected, curated, organized, and preserved for long term discovery, access, and use. Archivists often create finding aids that catalog groups of related items to assist with access. Archivists also employ highly specific preservation methods to ensure that archival assets are available for use well into the future. In planetary science, archives of physical samples are important and have widely varying archival needs. Most physical archives now provide digital access to some part of their collections and are actively working to digitize physical items as funding permits. (Copied from Section 2.1.3.1)

calibrated data: Data that has been corrected from instrumental effects and is provided in physical units, possibly edited, resampled, or reprojected.

community-produced tools: Those developed by stakeholders in the PDE in service to their own local interests or goals. Such tools may include software incidentally created in the course of a specific research project (which may or may not be financially supported), or by volunteer effort. Community-produced tools can be both high-impact and highly valuable but will, in general, serve the narrow interests or needs of the stakeholders producing them and their close collaborators.

CoreTrustSeal: CoreTrustSeal (https://coretrustseal.org) is an international certification for trustworthy archives and repositories. CoreTrustSeal is sponsored by the International Science Council (ISC; https://council.science). CoreTrustSeal certification is required for membership in the ISC-sponsored World Data System (https://www.worlddatasystem.org/).

data curation: The active and on-going management of data through its lifecycle. Curation activities enable data discovery and retrieval, maintain quality, add value, and provide for reuse.

data lifecycle: Information lifecycle management is a comprehensive approach to managing the flow of an information system's data and associated metadata from creation and initial storage to the time when it becomes obsolete and is either archived or destroyed.

data management: Enables the location, sharing, and reuse of data, and reduces the redundancy of data. Good data management reduces costs in terms of time and money. Data management involves all stages of the digital data lifecycle including capture, analysis, sharing, and preservation.

data producers: Personas who create data. Data creation may be accomplished through acquisition of low-level data from instruments; from field work; from laboratory work; by creating higher-level data products based on lower-level data; writing software; creating model outputs; etc.

data proposers: Intend to submit a proposal for NASA funding, with the proposal including the intent to archive data in either the PDS or other archive site. The proposal might be prepared as part of a mission proposal, or as an individual researcher’s response to a data analysis program call. A data proposer is looking for information on how to include archiving in their proposal to meet the requirements of data availability. Successful data proposers will later become data providers.

data providers: Researchers or research institutes that wish to make their data available through a public repository.

data user: Anyone (e.g. a researcher, student, educator, etc.) who is interested in obtaining, manipulating, and understanding data products. They intend to locate data products, find tools to work with those data products, and understand how those data products are formatted and structured.

derived data: Data products resulting from scientific analysis of calibrated data, possibly merging individual measurements (e.g. atmospheric profiles or mineralogical maps, from spectral measurements; feature catalogues, from images). Also known as secondary or higher-level data products.

FAIR principles: Guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets. The FAIR principles emphasize machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention).

higher-level data: See “derived data.”

information ecology: A system of people, practices, technologies, and values in a local environment (see Nardi and O’Day, 1999).

information model: A representation of concepts, relationships, constraints, rules, and operations to specify data semantics for a chosen domain of discourse. Specifically, the PDS Information Model (IM) is the representation that specifies PDS4.

infrastructure tools: Like power lines and bridges, infrastructure tools connect otherwise disparate elements of the PDE in a way that provides value to large cross-sections of planetary data users. Their cross-cutting nature also means that it is sometimes not clear that any single entity is primarily responsible (or “owns”) their development and maintenance, and it may therefore be necessary for NASA to directly initiate and support the development and ongoing maintenance of these systems. Examples of extant infrastructure tools within the PDE might include (but are not limited to) the following: Geospatial Information Systems for working with planetary mission data (perhaps specific capabilities added to QGIS); Search capabilities across planetary mission data (including but not limited to the PDS); Other software systems necessary to understand, process, calibrate, or use planetary data in general. E.g. SPICE, ISIS, GDAL, Ames Stereo Pipeline, Astropy, PlanetaryPy, etc.; The PDS information model and associated software systems.

International Heliophysics Data Environment Alliance (IHDEA): The goal of the International Heliophysics Data Environment Alliance (IHDEA) is to encourage the use of common standards and services in order to enable sharing of data and to enhance science. The idea of IHDEA has its origin in NASA's 2013-2022 Solar and Space Physics Decadal survey which identified that science is best served by “instant unfettered access to a wide array of datasets from distributed sources in a uniform, standardized format”.

International Planetary Data Alliance (IPDA): The International Planetary Data Alliance (IPDA) is a closely cooperating partnership to maintain the quality and performance of data from planetary research using instruments in space. Specific tasks include promoting the international exchange of high-quality scientific data, organized by a set of standards to facilitate data management. NASA's Planetary Data System is the de facto standard for archiving planetary data. Member organizations participate in both its Board and on specific projects related to building standards and interoperable systems.

International Virtual Observatory Alliance (IVOA): The International Virtual Observatory Alliance (IVOA) was formed in June 2002 with a mission to develop and agree upon the vital interoperability standards upon which the VO implementations are constructed. The IVOA is endorsed by the International Astronomical Union (IAU) as the accepted method for producing astronomical data-related standards. The IVOA includes a Solar System Interest Group (SSIG) that reviews IVOA standards in the scope of solar system sciences.

machine-readable data: Data in a data format that can be automatically read and processed by a computer, such as CSV, JSON, XML, etc.

Mapping and Planetary Spatial Infrastructure Team (MAPSIT): MAPSIT was established by NASA and the planetary science community in the fall of 2014, following recommendations from the Planetary Science Subcommittee of the Science Committee of the NASA Advisory council. Originally named the Cartography Research Assessment Group (CRAG), the MASPSIT name was adopted in the fall of 2015 to be more inclusive of all aspects of spatial data analysis and associated infrastructure. The team consists of all interested members of our community and has a Steering Committee which actively solicits input from the scientific community and reports its findings to NASA as requested.

metadata: Standardized descriptive information that goes with a data file (e.g. FITS keywords), or information in a registry entry that describes a resource such as a data service, or information that specifies the structure of a database - a list of its tables, their column names, and their UCDs. Some metadata are intended to provide references (provenance, access…), some to support searches (coverages, instrumental or observational parameters…).

name resolver: A service that associates all possible names or ID of an element, such as astronomical sources. Name resolvers permit databases to be searched independently from designation. For fixed sources, various catalogue IDs are associated with constant celestial coordinates (such as those from Simbad or NED). For solar system bodies, various designations can be associated with ephemeris. Name resolvers usually also provide some properties of the targets. The principle could be extended to elements other than sources, such as observatories, spacecraft lists, or authors.

NSPIRES: NASA Solicitation and Proposal Integrated Review and Evaluation System, a software system through which researchers submit proposals to ROSES.

persona: Personas are hypothetical narratives intended to represent the contextual data about real users (Hartson & Pyla 2012).

Planetary Data Ecosystem (PDE): 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.

Planetary Data Ecosystem (PDE) elements: Include repositories and communities relevant to the planning, acquisition, preservation, and use of the data collected by planetary space missions and research programs.

Planetary Data Ecosystem (PDE) tools: Tools or utilization resources are considered to be in scope of the PDE if their nearly exclusive use case is for the utilization of planetary data or if they contain a sub-component with a nearly exclusive purpose for utilization of planetary data. For example, the JMARS software suite would be considered in scope of the PDE because its sole purpose (barring excessive creativity) is for the utilization of planetary data. The ArcGIS and QGIS software systems are not exclusively designed for the utilization of planetary data, but subcomponents of these tools have been modified to exclusively support planetary data, (e.g. for non-terrestrial geoids), so these tools are considered in scope of the PDE. General-use software tools like text editors, operating systems, or image processing suites are critical to the use of planetary data, but they contain no functionality exclusive to the use of planetary data and are therefore out of scope.

Planetary Data System (PDS): A 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.

Planetary Software Organization: GitHub is an open-source location for both data and software. The Planetary Software Organization promotes open-source software in the planetary sciences by helping software creators and maintainers foster an open-source community. The Planetary Software Organization helps software creators grow and organize their community by providing a set of ready-to-use community guidelines, standards, and governance documents. The Planetary Software Organization also helps software maintainers manage the growth and evolution of their open-source software by providing mentorship and suggestions.

Planetary Spatial Data Infrastructure (PSDI): Planetary Spatial Data Infrastructure is an organizational framework, drawn largely from terrestrial spatial data infrastructures, to support the discovery and use of spatially enabled data by non-expert users. MAPSIT is largely focused on the development and use of PSDIs. It is a plan for obtaining and organizing data in a standardized way to make them discoverable, accessible and usable.

raw data: Data set corrected for telemetry errors and decommutated. Data are tagged with time and location of acquisition but have otherwise not been processed.

registry: Registries are catalogs. They do not hold data; Instead, they describe the location, access mode, and (to some extent) the content of tools, data sets and/or products. Registered products can include data files, label files, metadata and other schemas, dictionary definitions for objects and elements, services, and so on. Examining these products can give users insight into the structure, organization, and (to a lesser degree) the contents of related data sets. Registries are typically not curated, relying instead on data providers to submit and document content. (Copied from Section 2.1.3.4)

repository: Repositories, whether physical or digital, have less stringent standards for inclusion than archives, but are often curated to enable search, discovery, and reuse of their content. They may preserve, manage, and provide access to many types of materials in a variety of formats. Repositories vary widely in their standards and level of curation. A useful repository for the storage of science data must have sufficient control for digital material to be authentic, reliable, accessible, and usable on a continuing basis. Throughout this report, we will refer exclusively to digital repositories. (Copied from Section 2.1.3.3)

research data: The recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This “recorded” material excludes physical objects (e.g., laboratory samples). (Source: Office of Management and Budget (OMB) Circular A110)

ROSES: Research Opportunities in Space and Earth Science, a NASA grant program that funds research and analysis with NASA missions, programs, and data archives.

secondary data: See “derived data.”

service: Something on the internet which will actively do something for you, as opposed to being a passive repository of information. For example, an image service may have a large atlas of images, but also offers a way of submitting a query to get back a cut-out image from a particular piece of sky. (Source: IVOA glossary)

SPICE: An ancillary information system designed and implemented by NASA's Navigation and Ancillary Information Facility (NAIF). The letters making up “SPICE” stand for different types of information described in SPICE data: Spacecraft, ephemerides (Planet), Instrument, orientation (C-matrix), and Event information. SPICE is used throughout the life cycle of NASA planetary science missions to help scientists and engineers design missions, plan scientific observations, analyze science data, and conduct various engineering functions associated with flight projects.

tools: Software applications with generic capacities such as data visualization, manipulation, and analysis, and possibly on-line search in data repositories (including cross-matches), often restricted to some data types (e.g. images, spectra, tables, etc). VO tools or GIS are examples of these, in contrast with calibration/processing pipelines or computing environments. Tools often rely on protocols and standards, may be scriptable and may exchange data.

user experience (UX): A person's perceptions and responses that result from the use or anticipated use of a product, system or service. (Source: ISO 9241-210)

virtual observatory (VO): A collection of interoperating data archives and software tools that use the internet to form a scientific research environment in which astronomical research programs can be conducted.

I PDE Elements

When NASA invited the public to self-nominate for positions on the Planetary Data Ecosystem Independent Review Board, the call provided the following list of “PDE Knowledge Areas”:

- ADS
- AMMOS
- AstroMat
- Autoplot
- DAPs (Data Analysis Program, incl. Cassini DAP, Discovery DAP, New Frontiers DAP, Lunar DAP, and Mars DAP)
- IPDA
- JMars
- JPL HORIZONS
- JSC Curation online catalogs
- MAPSIT
- NASA Github
- NASA Planetary Github
- NSSDCA
- PDART (Planetary Data Archiving, Restoration, and Tools)
- PDS
- Planetary Geologic Mapping
- Planetary Photojournal
- PSIDA
- Quickmap
- RPIFs

This list provided the kernel of this Appendix. As described in Appendix C, the board significantly expanded the list of PDE elements. The list below still is not comprehensive. It is merely representative of the breadth and diversity of planetary data ecosystem elements. An element not being listed here does not imply that it is less important or valuable than listed elements.

I.1. Sources of New NASA Data (Other Than Missions)

I.1.1. DAPs (Data Analysis Program, incl. Cassini DAP, Discovery DAP, New Frontiers DAP, Lunar DAP, and Mars DAP):

Through Data Analysis Programs (DAPs), NASA funds researchers to enhance the scientific return of past missions and to increase the number and diversity of scientists involved in analysis of data. There are distinct Data Analysis Programs for Cassini (CDAP), Discovery missions (DDAP), New Frontiers missions (NFDAP), Lunar missions (LDAP), and Mars missions (MDAP). More are likely to be added in the future.

I.1.2. PDART (Planetary Data Archiving, Restoration, and Tools)

The Planetary Data Archiving, Restoration, and Tools (PDART) program solicits proposals to generate higher-level data products, archive and restore data sets or products, create or consolidate reference databases, generate new reference information, digitize data, and develop or validate software tools. The objective of this program element is to increase the amount and quality of digital information and data products available for planetary science research and exploration, and to produce tools that would enable or enhance future scientific investigations. Although it is expected that a small amount of data analysis, interpretation, or modeling may be performed to validate any generated products, this program element does not accept proposals in which the main focus is hypothesis-based planetary science. For all types of proposals, the products of selected proposals must be made available to the scientific community. (Source)

I.2. The Planetary Data System (PDS)

The Planetary Data System (PDS) is a 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. The PDS is a federation of nine teams geographically distributed around the U.S. Six are science discipline nodes, focusing on Atmospheres, Geosciences, Cartography and Imaging Sciences, Planetary Plasma Interactions, Ring-Moon Systems, and Small Bodies. There are two support nodes: the Engineering Node, and the Navigation and Ancillary Information Node. The ninth team is the project management group, which includes a project scientist function. (Source)

The PDS enables search and discovery of data via tools, services, and APIs for working with data following the PDS standards. Tools have been submitted from the broad PDS community and multiple institutions, including those from members of the International Planetary Data Alliance (IPDA). The Tools interface allows the user to search for and discover these tools. The interface also allows tool providers to submit their software for inclusion in the registry, but this registry may be out-of-date, and is only as good as the information submitted; it is not curated. Advanced, focused search tools are available from several PDS discipline nodes.

A Data Search function allows the user to search for data based on Target or Mission.

I.2.1. PDS Nodes

The activities supported at each node of the PDS are broad. Following is a brief description of each node and a few example resources hosted at each.

I.2.1.1. PDS Atmospheres Node (ATM)

The Planetary Atmospheres Discipline Node of the Planetary Data System (PDS) is responsible for the acquisition, preservation, and distribution of all non-imaging atmospheric data from all NASA planetary missions (excluding Earth observations). ATM also archives ground-based data acquired in support of the planetary missions, as well as preserving data from planetary analog laboratory and field measurements. (Source)

The Atmospheres Node has made special efforts to curate and make available to the PDS user planning materials, references, and useful summaries from several missions. These materials are relevant not only to the study of atmospheres, but to all other disciplines. For example, the Cassini Mission Archive includes mission overviews, objectives, results, and background pages on each of the science instruments, experiments, and mission science domains.

I.2.1.2. PDS Cartography and Imaging Science Node (CIS, formerly IMG)

The Cartography and Imaging Sciences Discipline Node (formerly known as the “Imaging Node,” IMG) of the PDS is the curator of NASA's primary digital image collections from past, present and future planetary missions. Imaging provides to the NASA planetary science community the digital image archives, ancillary data, sophisticated data search and retrieval tools, and cartographic and technical expertise necessary to develop and fully utilize the vast collection of digital planetary images of many terrestrial planetary bodies, including icy satellites. Imaging science expertise includes orbital and landed camera instrument development and data processing, data engineering and informatics, planetary remote sensing at UV to RADAR wavelengths, and cartographic and geospatial data analysis and product development.

Example resources at CIS include:

- Photojournal hosts publicly released images and captions from various solar system exploration programs. It includes all planetary science images released from the Jet Propulsion Laboratory, and some of those released by other NASA centers and NASA researchers.
- PILOT is an image data portal designed for searching through the raw image archives held by CIS. The catalog differs from other PDS catalogs because it has been improved by characterizing image geometry in great detail.

- The Image Atlas is a database-driven search tool that provides access to all image products held by CIS.
- Map Projection on the Web provides users with tools to crop and reproject planetary global cartographic data, perform other types of image analysis, and download the resulting image in the user’s choice of image formats.

I.2.1.3. PDS Engineering Node (EN)

The Engineering Node (EN) of the PDS provides technical support for all the discipline nodes. The Engineering Node manages the PDS Archiving Standard (PDS4) Information Model, context references, the PDS Central Registry, and provides website search and other technical support for the whole of PDS. Access to the Engineering Node website is limited to PDS staff.

I.2.1.4. PDS Geosciences Node (GEO)

The Geosciences Discipline Node archives and distributes digital data related to the study of the surfaces and interiors of terrestrial planetary bodies, in particular data acquired from the surfaces of these worlds. The Geosciences Node has developed several data search tools, including:

- The Analyst’s Notebook is an online tool for exploring planetary data from NASA Mars and lunar landed missions. The Notebook integrates sequence information, engineering and science data, and documentation. Currently, the Analyst’s Notebook includes exploration for Curiosity, InSight, Opportunity and Spirit, Phoenix, LCROSS, and Apollo missions.
- The Orbital Data Explorer (ODE) is a cross-mission and instrument query, search, display, and download tool for locating and retrieving PDS orbital science archives of Mars, Mercury, Venus, and Earth’s Moon.
- Spectral libraries (databases of laboratory spectra) contributed by the Mars Reconnaissance Orbiter CRISM instrument team and the Brown University RELAB laboratory, among others.

I.2.1.5. PDS Navigation and Ancillary Information Node Facility (NAIF)

NASA's Navigation and Ancillary Information Facility (NAIF) was established at the Jet Propulsion Laboratory to lead the design and implementation of the “SPICE” ancillary information system. The letters making up “SPICE” stand for different types of information described in SPICE data: Spacecraft, ephemerides (Planet), Instrument, orientation (C-matrix), and Event information. SPICE is used throughout the life cycle of NASA planetary science missions to help scientists and engineers design missions, plan scientific observations, analyze science data, and conduct various engineering functions associated with flight projects. In addition, as the “ancillary data node” of NASA's Planetary Data System (PDS), the NAIF Team leads the peer review of, and archives, the SPICE ancillary data products produced by NASA planetary flight projects. The NAIF node provides mechanisms for public access to these archived products. For example:

- The SPICE Toolkit consists of application program interfaces (APIs) that customers incorporate in their own application programs to read the SPICE ancillary data files and, using those data, compute derived observation such as altitude, latitude/longitude, and lighting angles, and to also determine various kinds of solar system events.
- The WebGeocalc (WGC) tool provides a web-based graphical user interface to many of the observation geometry computations available from the SPICE system. A WGC user can perform SPICE computations without the need to write a program; the user needs only a computer with a standard web browser.
- The Cosmographia Mission Visualization Tool is a SPICE-enhanced version of the open source visualization tool named Cosmographia. This interactive tool produces 3D visualizations of planet ephemerides, sizes and shapes; spacecraft trajectories and orientations; and instrument fields-of-view and footprints.

I.2.1.6. PDS Planetary Plasma Interactions Node (PPI)

The Planetary Plasma Interactions (PPI) Node of the Planetary Data System (PDS) archives and distributes digital data related to the study of the interaction between the solar wind and planetary winds with planetary magnetospheres, ionospheres and surfaces. The PPI Node is located at the Department of Earth, Planetary, and Space Sciences at the University of California, Los Angeles (UCLA).

I.2.1.7. PDS Ring-Moon Systems Node (RMS)

The Ring-Moon Systems Discipline Node (RMS) is devoted to archiving, cataloging, and distributing scientific data sets relevant to planetary ring systems, planetary moons, and the ways they interact. Because images of rings and moons provide important context to other types of measurements, there is substantial overlap among the holdings of the RMS and CIS nodes for outer-planets missions.

The Ring-Moon Systems Node is particularly dedicated to open science and to improving data accessibility and usability by a wide variety of potential users. They provide online tools for computing and visualizing the complex dynamics of ring-moon systems; share open-source software via Github; and produce higher-level versions of PDS-archived data sets from completed missions, such as a calibrated and geometrically corrected version of the Voyager data set. The Outer Planets Unified Search (OPUS) is the primary search tool for this node. The RMS Node generates geometric metadata above and beyond that archived by missions, enabling searches that include bodies fortuitously included in images targeted at other bodies, or searches by ring system geometry.

I.2.1.8. PDS Small Bodies Node (SBN)

The Small Bodies Node (SBN) specializes in archiving, cataloging, and distributing scientific data sets relevant to asteroids, comets, and interplanetary dust. Specifically, the SBN archives data from space missions (primarily NASA, but also from other national agencies) that target small bodies or that have collected small bodies data, laboratory and ground-based data collected in support of small bodies missions as well as contributed ground-based observations, and collected information about small bodies from published literature. All of the MPC's operating funds come from a NASA Near-Earth Object Observations program grant. There are several subnodes:

- The Comet Subnode is located at the University of Maryland, in College Park, Maryland. In addition to maintaining the combined archives of the SBN and supporting the SBN website, the Comet subnode collects, formats, verifies, and consults on data sets concerned with comet observations. The subnode also provides support for active comet missions and observing campaigns.
- The Asteroid/Dust Subnode is located at the Planetary Science Institute in Tucson, Arizona. The Asteroid subnode collects, formats, verifies and reviews ground-based and mission data pertaining to asteroids, transneptunian objects, small planetary satellites and interplanetary dust.
- The Minor Planet Center (MPC) is the single worldwide location for receipt and distribution of positional measurements of minor planets, comets, and outer irregular natural satellites of the major planets. The MPC is responsible for the identification, designation, and orbit computation for all of these objects. The MPC operates at the Smithsonian Astrophysical Observatory, under the auspices of Division F of the International Astronomical Union (IAU).

I.3. PDS-Equivalent Archives

Through a series of updates to the call for proposals to the PDART program, NASA has defined what it means for a data archive to be “PDS-equivalent.” Such archives must be:

- Independent: managed by someone other than the major data provider.
- Sustainable: managed for the long term (25 years at least).
- Open access: accessible to the public (lay and scientific) without pre-approval.
- Searchable.
- Citable.
- Preeminent: considered by its user community as the “standard” archive for the subfield.
- Standardized: data products must be provided in standardized formats and file types.

In addition, it is desirable for PDS-equivalent archives to provide peer review and documentation (user guides, calibration descriptions, etc.) for the data they hold.

The following are some (by no means all) examples of PDS-Equivalent archives.

I.3.1. Center for NEO Studies (CNEOS)

The JPL Center for NEO Studies (CNEOS) computes high-precision orbits for Near-Earth Objects (NEOs) in support of NASA's Planetary Defense Coordination Office. These orbit solutions are used to predict NEO close approaches to Earth and to produce comprehensive assessments of NEO impact probabilities over the next century.

CNEOS supports observers through the JPL HORIZONS high-precision ephemeris computation capability. HORIZONS is provided by the Solar System Dynamics (SSD) Group of the Jet Propulsion Laboratory.

CNEOS provides access to continually updated calculations of orbital parameters, close approaches, impact risks, discovery statistics, and mission designs via its website, with a data page for every NEO, and to user scripts through an API.

Parameters are archived in the JPL Small-Body DataBase (SBDB), which can be searched using a highly configurable filtering tool.

I.3.2. HIgh-resolution TRANsmission molecular absorption database (HITRAN)

HIgh-resolution TRANsmission molecular absorption database (HITRAN) is a compilation of spectroscopic parameters that a variety of computer codes use to predict and simulate the transmission and emission of light in the atmosphere. The database is a long-running project started by the Air Force Cambridge Research Laboratories (AFCRL) in the late 1960s in response to the need for detailed knowledge of the infrared properties of the atmosphere. The HITRAN compilation, and its associated database HITEMP (high-temperature spectroscopic absorption parameters), are developed and maintained at the Atomic and Molecular Physics Division, Harvard-Smithsonian Center for Astrophysics.

I.3.3. Infrared Science Archive (IRSA)

Caltech's Infrared Science Archive (IRSA) is chartered to curate the science products of infrared and submillimeter astrophysics missions led by NASA or with NASA involvement. Data of interest to planetary scientists have been collected by a number of these missions. IRSA serves the raw and pipeline-processed data associated with these missions, as well as enhanced data products contributed by the community. IRSA also curates data from large-area infrared surveys that are actively mined to yield discoveries in planetary science. In the future, IRSA will be the archive for data collected by the NEO Surveyor mission, implemented by the Planetary Defense Coordination Office.

IRSA's Guide for Solar System Observers introduces planetary scientists to relevant data sets and tools, including IRSA's “precovery” services, which allow researchers to mine data sets for serendipitous observations of near-earth objects (NEOs).

I.3.4. Mikulski Archive for Space Telescopes (MAST)

The Mikulski Archive for Space Telescopes (MAST) provides a variety of astronomical data sets focused on the optical, near-infrared, and infrared parts of the spectrum, including data from Hubble, TESS, Kepler, other NASA astrophysical spacecraft, shuttle astrophysics instruments, user-contributed High Level Science Products, and some ground-based observatories such as Pan-STARRS. In the 2020s MAST will host data from the James Webb Space Telescope and the Nancy Grace Roman Space Telescope. MAST offers a variety of search tools including search forms, APIs (including Virtual Observatory protocols), graphical exploration interfaces, a website for running SQL queries, and more.

I.3.5. NASA Planetary Geologic Mapping Program

The Astrogeology Science Center at the United States Geological Survey (USGS), based in Flagstaff, Arizona, coordinates the NASA Planetary Geologic Mapping Program. USGS map coordination is provided under the auspices of NASA's Planetary Cartography and Geologic Mapping Working Group and its Geologic Mapping Subcommittee. The maps produced through this program are published as part of the USGS Special Investigation Map (SIM) series. These maps are archived in the USGS publication repository that meets all Federal requirements for longevity, discoverability and accessibility.

I.3.6. NASA Space Science Data Coordinated Archive (NSSDCA)

The NASA Space Science Data Coordinated Archive (NSSDCA) serves as the permanent archive for NASA space science mission data. “Space science” includes astronomy and astrophysics, solar and space plasma physics, and planetary and lunar science. As the permanent (or “deep”) archive, NSSDCA serves a distinct purpose from NASA's discipline-specific space science “active archives,” such as the PDS. NSSDCA’s collections include hard-copy data and documentation that has not yet been digitized and may not be available elsewhere, including information from NASA’s earliest missions. NSSDCA also serves as NASA's permanent archive for space physics mission data.

Web-based services allow the NSSDCA to support the general public with information about spacecraft, science instruments, solar system object physical and orbital data, and access to digital versions of selected imagery. NSSDCA’s online services also provide information about data archived at NSSDCA (and, in some cases, other facilities) that is not available online. NSSDCA is part of the Solar System Exploration Data Services Office (SSEDSO) in the Solar System Exploration Division at NASA's Goddard Space Flight Center. NSSDCA is sponsored by the Heliophysics Division of NASA's Science Mission Directorate. NSSDCA acts in concert with various NASA discipline data systems in providing certain data and services.

I.3.7. NASA Science and Technical Information Program

The NASA Science and Technical Information program hosts the Research Access Initiative, part of the agency’s framework for increasing public access to scientific publications and digital scientific data. The STI site hosts access to the PubSpace repository, where all NASA-funded authors and co-authors (both civil servant and non-civil servant) are required to deposit copies of peer-reviewed scientific publications and associated data. That data is shared via the National Institutes of Health’s (NIH) PubMed Central (PMC) system.

I.3.8. Space Physics Data Facility (SPDF)

The Space Physics Data Facility (SPDF) is the NASA active and permanent archive for non-solar heliophysics data (solar data at SDAC), per the NASA Heliophysics Science Data Management Policy. SPDF is a project of the Heliophysics Science Division (HSD) at NASA's Goddard Space Flight Center. SPDF also provides multi-project, cross-disciplinary access to data to enable correlative and collaborative research across discipline and mission boundaries with present and past missions. SPDF maintains the SSCweb database of spacecraft orbits, the OMNIweb cross-normalized database, and the Common Data Format (CDF) self-describing science data format and associated software. Also the CDAWeb, the Coordinated Data Analysis (Workshop) Web, supports interactive plotting of variables from multiple instruments on multiple investigations simultaneously on arbitrary, user-defined time-scales. It also supports data retrieval in both CDF or ASCII format. Considered PDS-equivalent.

I.4. Physical Facilities

Note that most physical facilities provide at least some of their data collections in digital format.

I.4.1. NASA Center Archives

Most (though not all) NASA centers have physical archives. An example NASA Center Archives is the Jet Propulsion Laboratory Archives. Different from libraries, the primary mission of archives is to document the rich organizational, mission, and cultural histories of the institution by identifying, collecting, preserving, and making available primary source materials that have long-term value for research by users at the institution and the wider public. Archives provide research assistance, including assistance with locating historical records and answering questions related to the institution’s projects and activities. Some materials are cleared for public release. Physical visits are by appointment and application only.

I.4.2. NASA Regional Planetary Image Facilities (RPIFs)

The NASA Regional Planetary Image Facilities were an international system of planetary image libraries established in 1977. In 2019, the NASA History Office reviewed all of the unique content of the nine US RPIFs, and a report was provided in April 2020 of high priority items requested by NASA to be digitized. At the same time, funding for the RPIFs was discontinued. Many, though not all, of the home institutions have plans to maintain at least some of their collections and resources for public access.

These facilities maintain photographic and digital data as well as mission documentation and cartographic data. Each facility's general holding contains images and maps of planets and their satellites taken by solar system exploration spacecraft. These planetary image facilities are open to the public. Although there is some overlap among their collections, most RPIF collections vary from institution to institution, reflecting the expertise of researchers at the host institutions. In fact, each RPIF node hosts its own unique collection of data accumulated by planetary scientists in their region, as well as rescued mission and research data sets and literature) from departed investigators or closed centers. Experienced staff can assist scientists, educators, students, media, and the public in ordering materials for their own use.

The RPIF nodes have worked to digitize many of their unique document holdings and place them online at their individual institution websites as funding allows, transforming themselves from libraries of hard-copy material to planetary data utilization service and training centers.

As of 2020, the following RPIF nodes were active:

- The JPL Regional Planetary Image Facility, Jet Propulsion Laboratory, Caltech, Pasadena, CA
- The LPI Regional Planetary Image Facility, Lunar and Planetary Institute, Houston, TX
- The NASA/USGS Astrogeology Regional Planetary Information Facility (RPIF), Flagstaff, AZ
- The Northeast Regional Planetary Data Center, Brown University, Providence, RI
- The Pacific Regional Planetary Data Center, Hawaiʻi Institute of Geophysics and Planetology, University of Hawaiʻi at Manoa
- The Ronald Greeley Center for Planetary Studies, Arizona State University, Tempe, AZ
- The Smithsonian Regional Planetary Image Facility, Center for Earth and Planetary Studies (CEPS) at the Smithsonian National Air and Space Museum, Washington, DC
- The Space Imagery Center, Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ
- The Spacecraft Planetary Image Facility, Cornell University, Ithaca, NY
- And 7 international locations

I.4.3. Spectroscopy Laboratories

There are many spectroscopy laboratories across the country that are available for in-person use by NASA-funded researchers and which host digital databases of reference spectra for materials relevant to planetary science studies. Following is a non-exhaustive list:

- RELAB (Keck/NASA Reflectance Experiment Laboratory), a multi-user spectroscopy facility hosted by Brown University in Providence, Rhode Island
- The Thermal Infrared Spectral Laboratory and Mineral Library at Arizona State University in Tempe, Arizona includes observations of terrestrial rock and mineral samples for comparison with the spectra of Mars returned by thermal emission spectrometers and also for interpreting remote sensing data collected with Earth as the target. https://speclib.asu.edu/
- The USGS Spectroscopy Laboratory hosts a spectral library of thousands of lab measurements to facilitate laboratory and field spectroscopy and remote sensing for identifying and mapping minerals, vegetation, and manufactured materials. https://www.usgs.gov/labs/spec-lab/capabilities/spectral-library

I.4.4. The Astromaterials Acquisition and Curation Office at Johnson Space Center (JSC)

The Astromaterials Acquisition and Curation Office at Johnson Space Center (JSC), part of the Astromaterials Research and Exploration Science (ARES) Division of NASA, is responsible for the curation of extraterrestrial samples from NASA's past and future sample return missions. Their mission includes the documentation, preservation and preparation of samples from the Moon, asteroids, comets, solar wind, and the planet Mars. Their highest priority is to secure the future availability of these samples for the worldwide scientific community. This office sponsors two programs:

- Lunar and Meteorite disk that make samples available for use (2 weeks loan period) by K-12 educational institutions.
- Linear Petrographic Thin Section Designed for colleges and universities offering a curriculum in the geosciences. The petrographic thin section package is intended for use in college and university courses in petrology and microscopic petrography for advanced geology students.

I.5. Online Repositories, Registries, and Portals

Many facilities provide search and browse access to PDS as well as non-PDS data. Data and metadata vary widely in quality, discoverability, accessibility, interoperability, and permanence.

I.5.1. Astromaterials Data Repository (AstroRepo or AstroMat)

The Astromaterials Data Repository is based on EarthChem Library. AstroMat invites contributions of a broad range of extraterrestrial data, including but not limited to: compositional data for samples of lunar rocks, meteorites, minerals, melt and fluid inclusions, and more; geochemical synthesis data sets; geochronological data; petrographic descriptions of samples; kinetic data from geochemical and petrological experiments. They provide data templates to assist the efficiency of data set publication, and require data sets to meet guidelines for inclusion.

I.5.2. Astrophysics Data System (ADS)

The SAO/NASA Astrophysics Data System (ADS) is a digital library portal for researchers in astronomy and physics, operated by the Smithsonian Astrophysical Observatory (SAO) under a NASA grant. The ADS maintains three bibliographic databases containing more than 13 million records covering publications in Astronomy and Astrophysics, Physics, and the arXiv e-prints. Abstracts and full-text of major astronomy and physics publications are indexed and searchable. ADS tracks citations and usage of its records to provide advanced discovery and evaluation capabilities. Integrated in its databases, the ADS provides access and pointers to a wealth of external resources, including electronic articles available from publisher's websites, astronomical object information, data catalogs and data sets hosted by external archives.

I.5.3. Institutional Repositories

Many research institutions have established repositories to provide public access to data from publications by institution scientists -- a local solution to the need for repositories for science data. A few examples:

- Center for Earth and Planetary Studies (CEPS)
- JHU Applied Physics Laboratory Data Repository
- USRA Houston Repository

NASA Astrobiology Environments Database (AHED)

The NASA Astrobiology Environments Database is a new NASA-funded database, with a pilot study planned in early 2021 and full release early 2022.The Astrobiology Habitable Environments Database (AHED) is a long-term repository and productivity platform for the storage, discovery and analysis of data relevant to the field of astrobiology. AHED is built around an astrobiology specific standardized metadata framework (called ARMS – Astrobiology Resource Metadata Standard). The AHED Portal provides a web-based home to the project allowing new and returning users to create new ARMS compliant data sets and search for relevant data sets using a range of search tools designed around the needs of astrobiologists.

I.5.4. NASA Exoplanet Archive

The NASA Exoplanet Archive is an online astronomical exoplanet and stellar catalog and data service that collates and cross-correlates astronomical data and information on exoplanets and their host stars, and provides tools to work with these data. The archive is dedicated to collecting and serving important public data sets involved in the search for and characterization of extrasolar planets and their host stars. These data include stellar parameters (such as positions, magnitudes, and temperatures), exoplanet parameters (such as masses and orbital parameters) and discovery/characterization data (such as published radial velocity curves, photometric light curves, images, and spectra).

I.5.5. NASA Open Data and Software Portal

data.nasa.gov is NASA's clearinghouse site for open-data provided to the public, listing thousands of data sets. Some listed data sets link to resources available from other NASA data archives and repositories, while other data sets only exist on data.nasa.gov.

I.5.6. Scientific Journals

Various journals that publish planetary science results act as informal archives by providing access to data underlying published figures, machine-readable data tables, etc. Many members within the community use journals as their primary long-term archive, but the sufficiency of these archives can vary between journals.

I.6. NASA-Funded Public Communication

NASA’s public communication efforts are designed to serve three primary audiences: the general public, educators, and news media. All three are served through NASA’s websites at nasa.gov. Products include press releases, captioned images and videos, mission blogs, press kits, fact sheets, explainers, classroom resources, webinars for educators, lesson plans, laboratory activities, museum and planetarium resources, “museum in a box”, and much more. The following is only a partial list with examples of the kinds of public engagement activities that NASA supports.

I.6.1. Active Mission Websites

While NASA missions are active, they are funded at the mission level by NASA to do public engagement, sharing information primarily via mission websites but also a variety of other pathways. Each mission takes a unique approach to public engagement. Some are hosted at NASA centers, others at the home institution of a principal investigator. Once funding for a mission has ended, so does active public engagement. Mission websites are often left in their final state as a legacy of the mission and are not guaranteed to remain live or functional over time.

I.6.2. Raw Images Websites for Active NASA Missions

Many NASA planetary missions have an automated pipeline that converts image data to accessible formats (e.g. JPEG or PNG) and serves them on the Web for the public to follow daily mission activities. The availability of these archives has inspired the development of fan communities surrounding the missions. As of March 2021 currently active planetary mission raw images websites include Curiosity, InSight, Juno, and Perseverance. The Perseverance raw images website includes metadata in JSON format that has inspired a community of volunteer software developers to create unique image browse tools. Many Heliophysics and Earth science missions also share images immediately with the public.

Past planetary missions that shared raw images in this way include the Mars Exploration Rovers, Cassini, Phoenix, Dawn, and New Horizons. A few of these, notably Juno but also Cassini, invited visitors to submit processed versions of raw images to the websites to be shared with the public.

I.6.3. Instrument-Specific Websites

Many science instruments, especially those on flagship missions, perform public engagement and outreach out of their home institutions, providing access to science data as well as derived products and interpreted data on team websites. Because mission data management plans do not cover public engagement activities, funding and support for these websites usually cease with the end of a mission, though the websites often remain available for some time after end of mission. A few examples of these are listed below.

- High-Resolution Imaging Science Experiment (HiRISE) on Mars Reconnaissance Orbiter: HiRISE science data is available through the PDS. The HiRISE website also provides captioned images, slides, audio, and video (HiClips), intended for anyone, especially educators, to use as part a discussion, presentation, or class units about Mars. The BeautifulMars Project coordinates worldwide volunteer translators of these materials, allowing HiRISE to offer educational resources in 28 languages, which is the most of any active NASA mission.
- Lunar Reconnaissance Orbiter Camera: The Lunar Reconnaissance Orbiter Camera (LROC) Science Operations Center (SOC) team website provides a variety of browse and search tools for PDS-archived LROC data through a web portal. One such tool is Lunaserv Global Explorer, a Web Map Service (WMS) implementation for the Moon. The LROC team website also provides many “featured images” with captions explaining science interpretations and exploration history, browsable through the proprietary QuickMap spatial interface.

I.6.4. NASA Photo Galleries

See https://www.nasa.gov/multimedia/imagegallery/index.html for a long list of scattered photo galleries hosted at a variety of locations, some of them on NASA websites, others on commercial websites like Flickr.

I.6.5. News Media Information

Press releases, press kits, photos, videos, etc. are posted on mission websites and NASA’s main website. Many of these materials are selected for preservation in National Archives and/or Center archives. Materials older than the Internet are often not available on the Internet and therefore not readily available to the public.

I.6.6. Science Nuggets

Each of the Assessment Groups that provides community advice to NASA produces "science nuggets" summarizing recent developments in planetary science (e.g. the Small Bodies Assessment Group at https://www.lpi.usra.edu/sbag/science/ .)

I.6.7. Treks

Solar System Treks are online, browser-based portals that allow you to visualize, explore, and analyze the surfaces of other worlds using real data returned from a growing fleet of spacecraft. You can view the worlds through the eyes of many different instruments, pilot real-time 3D flyovers above mountains and into craters, and conduct measurements of surface features. The portals provide exciting capabilities for mission planning, planetary science, and public outreach.

I.7. NASA-funded K-12 Education Portals

Educational resources provide planetary information, press releases, derived data products, and raw data for educators to bolster state- and district-wide curricula. Next Generation Science Standards (NGSS) provide rubrics for a phased approach to teaching science at all levels and can utilize introductory space and Earth science materials, data to support graphing and plot reading skills, and raw data for classroom demonstrations and independent research (science fairs and advanced college credit) by the time students begin to enter undergraduate college programs.

I.7.1. Challenger Centers for Space Science Education

Created in the aftermath of the Challenger shuttle accident, the crew's families came together to create the Challenger Center for Space Science Education to carry on the legacy of their loved ones by continuing their educational mission. The Challenger Learning Centers are a global network of space-themed simulated learning centers that use role-playing strategies to help students bring their classroom studies to life, cultivating skills for future success, such as problem solving, critical thinking, communication and teamwork.

A Non-profit educational organization for engaging students and teachers in dynamic, hands-on exploration and discovery opportunities that strengthen knowledge in science, technology, engineering, and mathematics (STEM), inspiring students to pursue careers in these fields, and providing an outlet to learn and apply important life skills. There are ~40 centers across the United States, one in Canada, and one in South Korea.

Recent activities have included student-led simulated exploration missions to the Moon, Mars, and Europa that use NASA planetary data from recent missions and revolve around Human Exploration and Operations Mission Directorate (HEOMD) goals.

I.7.2. NASA Kids’ Club

NASA Kids’ Club is a safe place for children to play as they learn about NASA and its missions. The site provides games at various skill levels for children pre-K through grade 4. The games support national educational standards in STEM. NASA Kids’ Club Picture Show is an image gallery of some of NASA’s coolest and most interesting pictures. The site is focused on children but is used by parents and teachers to provide a gateway into space science for pre-K through grade 4.

I.7.3. NASA Science Space Place

Launched in 1998, NASA Space Place's mission is to inspire and enrich upper-elementary-aged kids' learning of space and Earth science online through fun games, hands-on activities, informative articles and engaging short videos. With material in both English and Spanish and resources for parents and teachers, NASA Space Place has something for everyone.

I.7.4. NASA STEM Engagement

Resource site for students and educators, with STEM resources organized by grade level and links to NASA Education resources.

I.8. Data Standards

I.8.1. PDS Data standards

When the NASA PDS began to archive data, relevant standards were generally absent. As such, the PDS undertook the Herculean task of establishing the first widely used standards for planetary data. While initially focused on NASA missions, many other space agencies leveraged PDS standards for their own mission products. This became more formal through the Consultative Committee for Space Data Systems (CCSDS) where NASA is one of the 11 member agencies, along with 29 observer agencies and has become a coordinated international effort through the International Planetary Data Alliance. Along with becoming an internationally accepted standard, the most recent PDS standard (PDS4) makes significant attempts to be compatible with other data standards, especially those used for terrestrial remote sensing and astronomy. The PDS Data Standard Reference Guide can be found at: PDS Data Standard Reference Guide

I.8.2. Examples of other relevant data standards

FGDC (Federal Geographic Data Committee) https://www.fgdc.gov/. The FGDC standards were developed in order to assure that geospatial data produced at the Federal, State, and local levels were all interoperable. Federal agencies are required to follow FGDC standards and many other organizations follow suit. In recent years there has been a strong push to align FGDC standards with those of the Open Geospatial Consortium (OGC) and ISO. The PDS3 format did not conform to FGDC but PDS4's geospatial metadata is based on FGDC.

ISO (International Organization for Standardization) https://www.iso.org/standards.html. As its name indicates, ISO exists to establish standards that are accepted internationally for a full range of products, ranging from how ropes are spooled to the definition of JPEG image standards. It is an independent, non-governmental organization, but its members include 165 national standards bodies such as the American National Standards Institute (ANSI). Conforming to ISO standards is generally considered the most reliable way to assure compatibility across national borders. Most planetary data, including PDS products, are not directly following ISO standards. However, as PDS follows FGDC other standards that are moving into alignment with ISO, the standards are drawing closer over time.

International Astronomical Union (IAU) https://www.iau.org/. The IAU is the internationally recognized body for setting the names of astronomical bodies and related information (such as coordinate systems). Given the range of unusual objects in the cosmos, the IAU often provides room for some flexibility by publishing “recommendations” rather than strict dictates. Overall, there is a strong emphasis on providing consistency between past published work and new studies but major changes that overturn decades of precedence are not unknown. The PDS has a long-standing requirement that PDS-archived data must conform to IAU standards (including recommendations).

I.9. Software

The following is a very incomplete list, intended to provide a few examples of the variety of software in the PDE.

I.9.1. Advanced Multi-Mission Operations System (AMMOS)

The Advanced Multi-Mission Operations System, or AMMOS, is a set of mission operations and data processing capabilities for robotic missions through an ‘Ops in a Box’ approach. AMMOS is a low-cost, highly reliable system utilized by more than 50 missions, including planetary exploration, deep space, earth science, heliophysics, and astrophysics, by NASA, ESA, industry, and academia.

I.9.2. Autoplot

An interactive browser for data on the web; give it a URL or the name of a file on your computer and it tries to create a sensible plot of the contents in the file. Autoplot was developed to allow quick and interactive browsing of data and metadata files that are often encountered on the web. It was developed under the NASA Virtual Observatories for Heliophysics program in a collaborative effort among several institutions, including support or code contributions from ViRBO, VMO, RBSP-ECT, and the Radio and Plasma Wave Group at The University of Iowa.

I.9.3. Integrated System for Imagers and Spectrometers (ISIS)

https://isis.astrogeology.usgs.gov/

https://github.com/USGS-Astrogeology/ISIS3

https://isis.astrogeology.usgs.gov/

I.9.4. JMars

https://jmars.asu.edu/

I.9.5. Small Body Mapping Tool

Spacecraft missions return massive amounts of valuable data, but those data can be hard to access, analyze, and interpret. Asteroids, comets, and small moons present additional challenges: the irregular shapes of these bodies are ill-suited for two-dimensional projections. The Small Body Mapping Tool (SBMT) addresses these challenges. The SBMT is an interactive tool that allows users to visualize and manipulate small body shape models in three dimensions. The Tool enables quick and easy searches for spacecraft data of a variety of small bodies. Once selected, data can be projected directly onto the shape models, and built-in analysis and mapping capabilities facilitate scientific investigations.