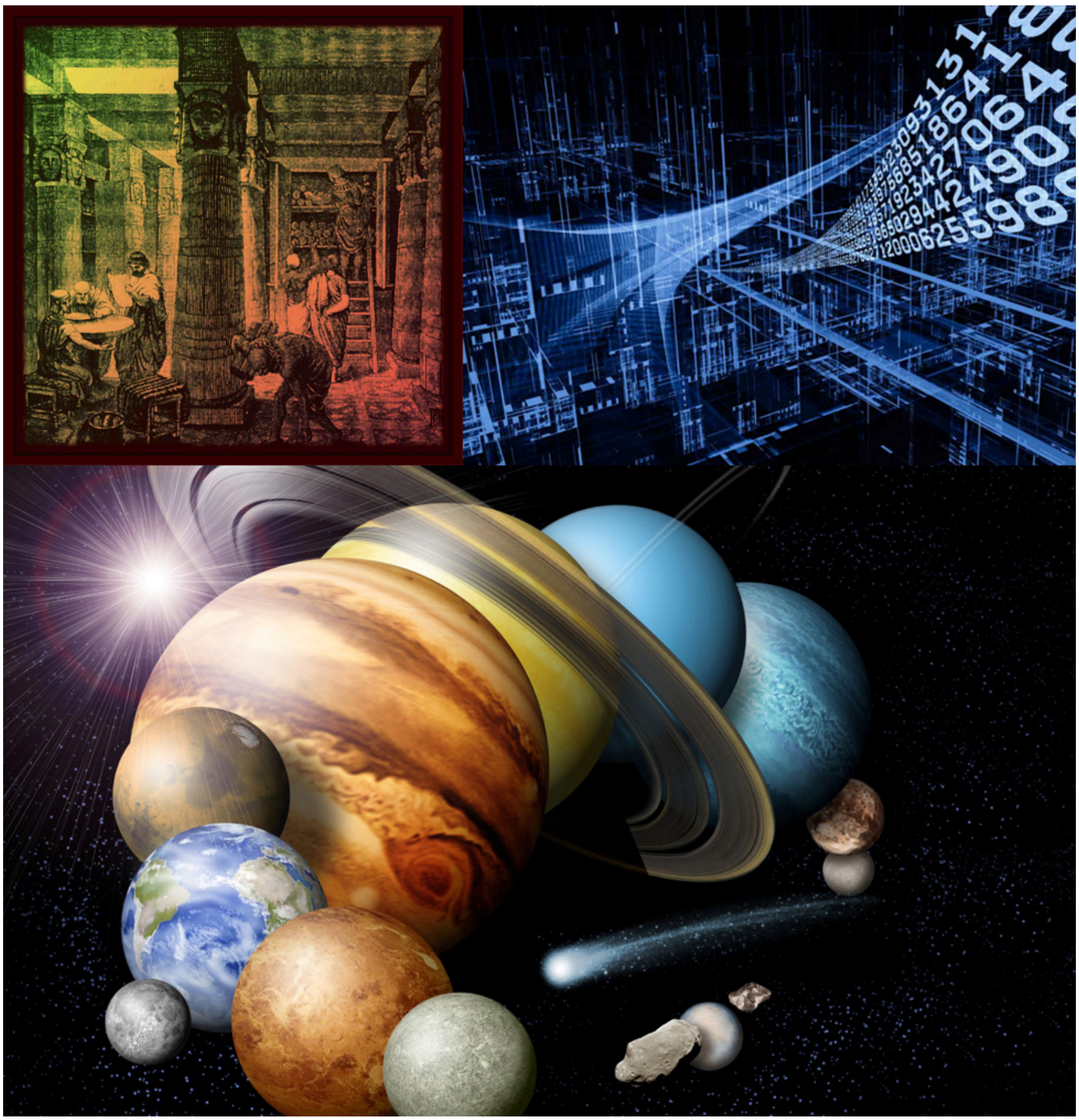
National Aeronautics and Space Administration

Planetary Data System Roadmap Study for 2017 – 2026



www.nasa.gov

Dedicated to the memory of Dr. Michael F. A'Hearn

Photo credit: Image Courtesy of Ball Aerospace & Technologies Corp.

Mike A'Hearn (c. 2004) with the copper "cratering mass" that capped the Deep Impact mission impactor spacecraft. The impact excavated a crater approximately 100-meters wide and 30-meters deep in the nucleus of comet 9P/Tempel 1.

Table of Contents

Photo credit: NASA/Wendy Dolci

Members of the NASA Astrobiology Institute (NAI) Executive Council view an image of Mars on the NASA hyperwall at Ames Research Center.

Executive Summary and Findings

Executive Summary and Potential PDS Actions for the Next Decade

The Planetary Data System (PDS) has made great strides in its almost three decades of existence, with the tide now turned against what was perceived as a threat of loss of planetary data in 1982 [1]. The evolving background of international archiving standards led first to the implementation of the PDS archive based upon the PDS3 standards, and, during the last decade to change from PDS3 to PDS4. This evolution has been driven primarily by the same motivations that set PDS into motion from the beginning: user needs and expectations (two findings), data discoverability (three findings), and data usability (three findings). These functions are supported by tools and file formats (two findings), and online processing and analysis (one finding). Changes in trends, as well as the potential for disruptive technologies, must be accounted for. These changes are reflected by increases in data volume, variety, complexity, and number of data providers (two findings), possible additions of laboratory data and physical samples (two findings), and potential enhancements of PDS structure and governance (three findings).

After introducing the PDS (Chapter 1) and laying out its characteristics (Chapter 2), we provide details on challenges facing the PDS (Chapter 3), detailed findings and suggested remediations to which these lead (Chapter 4), and conclusions and a summary of what the future may portend (Chapter 5). Exact responses to the various challenges depend upon external changes in technology (opportunities driven by new commercial hardware as well as computer security challenges), data production by robotic planetary missions (both domestic and international), and data needs and requirements from an evolving stakeholder base.

As a complex evolving system, the PDS must constantly respond to new pressures and opportunities. In turn, these generate user needs and expectations, especially with respect to PDS stakeholders (Finding I) and can lead to mismatches between those tasks PDS is equipped to take on and expectations of the stakeholders (Finding II). Key to supporting this system is having an agile and scalable architecture, e.g., the PDS4 information model, and software services and tools, that is adaptable to differing and competing priorities and needs. While competing priorities are a reality for a distributed, internationally-adopted system, PDS4 can evolve and be extended over time as priorities are defined and as funding is available.

Data stored by the PDS must be discoverable, i.e., be easily and efficiently searchable (Finding III) as well as interoperable with other archives (Finding IV) and easily citable to enable use by multiple researchers (Finding V). Key to this response is the modernization of the associated metadata, enabled by the PDS4 information model and its implementation (Finding VI) and how the model has enabled better access to the data (Finding VII).

A related, important, and large job is the migration of appropriate PDS3 archives into PDS4, so as to modernize all the relevant metadata associated with the current holdings. This task is particularly time-critical for missions still operating that were “grandfathered” into PDS3 requirements. To the extent that some unique metadata required for PDS4 resides with currently knowledgeable and active scientists, there is a need to tap into that knowledge-base sooner rather than later. This is especially true for Solar System targets that may not be revisited for a decade or more—e.g., Mercury, the Saturn system, and the Pluto system—though preserving data is important for all Solar System targets.

Scientific progress rarely follows a deterministic path. There is always the possibility of data once thought irrelevant becoming key to further advances (e.g., the Rosetta Stone). Maximizing the use of such data would be enabled by eventually migrating all PDS holdings to PDS4 as best as possible. In a world of finite resources, the responsible manager must make choices as to priorities of which data to migrate sooner than others; but, if resources were available, the conversion of the entire holdings would provide the best discoverability "insurance policy" (Finding III). To enable efficient use of the PDS4 potential, additional broad documentation and training are needed to supplement the already extant technical documentation (Finding VIII).

Efficient and effective use of PDS4 also requires the presence of adequate translation tools (Finding IX) and associated other tools. Common tool development is essential for PDS4 use by both the nodes and the user community. While there is continuing effort to create more tools, additional resources could be used to expedite such development (see, e.g., Findings IX and XII). Software archiving *per se* is not included in the PDS4 information model and creates its own challenges not covered currently by the PDS (Finding X).

Implementation of PDS tasks in an efficient manner requires staying abreast of changes in information technology, especially in areas related to dealing with extensive diversity in types of data (Finding XI). This also relates to foreseen increases in data volume, variety, complexity, and in number of data providers (Findings XII and XIII). The expansion of the PDS to include more data related to laboratory samples may be warranted but would also introduce new challenges (Findings XIV and XV).

From its inception in 1989, the PDS has undergone some significant structural changes while other aspects have remained remarkably stable. PDS began with a centralized structure with a Central Node, Discipline Nodes, and supplementary Data Nodes [2]. As the system evolved, Data Nodes continue to be employed on an as needed basis, while Discipline Nodes remain as a fixed part of the system [3]. By 2006 the Central Node functions had been broken into an Engineering (Support) Node at JPL and a much reduced Project Office at Goddard Space Flight Center (Appendix C and [4]). During this same period, the structure of the Discipline Nodes remained fairly stable (Finding XVI). The Project Office is charged to "manage funding and budgets and establish common policies across the PDS," but a lack of resources has limited the manner in which some of these activities are being carried out (Findings XVII, XVIII, and XIX).

Conceived to preclude irretrievable loss of robotic science data from throughout the solar system, the PDS can only be judged as an incredible success story. Maintenance of that story against new challenges and rising expectations is important and possible, but only with resources in excess of those now maintaining the current system.

Findings

The Roadmap Study Team (RST) is not a committee operating under the Federal Advisory Committee Act (FACA)[1]. Although the RST has presented "Findings" as statements of fact, the RST is not empowered to present recommendations as such. Nevertheless, there has been an attempt to present conceptual "remediations" that could respond to some of the 19 Findings reported herein.

These remediations are meant to provide specific plausible examples of mitigating or corrective actions that might be undertaken in response to problems or potential problems identified in the Findings. Undertaking these actions would necessarily require support, typically in the form of additional funding.

User Needs and Expectations

Finding I: PDS Stakeholders. While all PDS stakeholders are recognized as valuable, the prioritization of stakeholder interests and the impact those interests should have on PDS policy, design, and resource allocation are unclear.

Finding II: Managing Expectations of PDS Usability. There is a mismatch between the services and functions PDS is equipped to provide and the very high expectations of its users and NASA management.

Data Discoverability

Finding III: Data Discoverability. 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.

Finding IV: Integration with Other Archives. The PDS serves as the model for other national space-mission data archives in ensuring future universal accessibility and searchability. The PDS is uniquely poised to lead efforts to make national and global archives interoperable.

Finding V: Citation of Data Sets. PDS is actively involved in addressing the data citation issue, and is well-positioned to provide the essential links in the chain needed to enable clear, direct referencing of PDS products; but it cannot itself change the habits and attitudes of authors, referees, and journal editors when it comes to including data set references in publications.

Data Usability

Finding VI: Modernizing Metadata. The accessibility and discoverability through the PDS4 metadata registry is a cornerstone to the future of community interaction with the PDS as a coherent storehouse of data. Legacy data archived in PDS3 format (the vast majority of PDS holdings) often lack metadata sufficient to enable discovery and accessibility commensurate with PDS4.

[1] https://www.gsa.gov/portal/content/104514

Finding VII: Access to Data. The PDS does an excellent job of providing access to its data holdings and is on track to increase such access. The latter is enabled by the PDS4 uniform metadata standard.

Finding VIII: Documentation and Training. 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.

Tools and File Formats

Finding IX: PDS File Formats and Translation Software. There is a need for more translation programs that transform data from the PDS4 archive file formats to more usable analysis-ready formats.

Finding X: Archiving Software. The PDS is not an appropriate archive or repository for software.

Online Processing and Analysis

Finding XI: Information Technology. The PDS has been and continues to be proactive in investigating information technology and adopting best practices.

Increases in Data Volume, Variety, Complexity, and Number of Data Providers

Finding XII: Potential Impact of ROSES Archiving Requirements. It is a matter of concern as to whether the PDS nodes will have the resources to serve the data archiving requirements of individual ROSES investigations.

Finding XIII: Higher-Order Data Products. Higher-order products produced by mission teams beyond what is in their original data management plans are extremely valuable additions to the archive but are not always included due to lack of resources needed by missions to complete the archiving process.

Laboratory Data and Physical Samples

Finding XIV: Astromaterials Data I. A large amount of data from laboratory analyses of samples obtained by NASA missions is not archived and is in danger of loss. Astromaterials data today are primarily stored on short-lived media, in private holdings, and with PI-dependent documentation.

Finding XV: Astromaterials Data II. A large amount of data from laboratory analyses of meteorites and cosmic dust is not archived and is in danger of loss.

PDS Structure and Governance

Finding XVI: Node Organization. PDS funding levels, combined with the lack of opportunity to propose new nodes separate from the re-compete activity for existing nodes, has had the effect of strongly discouraging the establishment of new nodes or otherwise allowing the PDS organization to grow to keep pace with development and expansion of Planetary Science disciplines and technology.

Finding XVII: Transparency. The use and application of PDS4 standards and development of third-party support for PDS4 metadata and formats is hindered by a lack of transparency in the PDS development process.

Finding XVIII: PDS Governance. NASA management has not settled the question of how PDS fits into current NASA governance structures. PDS has a minimal Project Office, which lacks resources for providing detailed cross-discipline reports, studies, and guidance as there are within other NASA SMD data activities, which would put a more unified public face on the PDS and support other activities provided for in the current NASA governance model.

Finding XIX: Timing of the Next PDS Roadmap Study. This Roadmap Study was initiated in the year immediately following a recompetition of the PDS Nodes, and will be completed at least three years (and perhaps longer) before the next recompetition, which limits the impact of a Roadmap Study activity on shaping the work of the PDS.

1 Introduction

Since the launch of Sputnik 1 in October 1957, some 10,000 objects—including spacecraft, upper stages, and debris—have been launched to or beyond the orbit of the Earth [5, 6]. The NASA Space Science Data Coordinated Archive (NSSDCA) references 7,583 satellites, systems, and programs.[2] Of these, 1,843 satellites have carried out significant science missions over 60 years. The NSSDCA classifies missions by scientific discipline, although the same mission may address more than one discipline; there are 319 Astronomy missions, 994 in Earth Science, 316 in Planetary Science, 202 in Solar Physics, and 656 in Space Physics (the numbers sum to more than 1,843 due to multi-discipline overlaps). Most of these scientific satellites are from the United States and are sponsored, at least in part, by the National Aeronautics and Space Administration (NASA). From Pioneer 1 through OSIRIS-REx, there have been 117 successful such robotic missions. Scientific data from NASA's missions is distributed among 16 formal NASA Archives plus 22 Guest Observer Facilities and Science Centers.[3] Some of these entities are divided into more specialized organizations for supporting archiving and data usage, such as the 12 Distributed Active Archive Centers (DAACs)[4] of the Earth Observing System Data and Information System (EOSDIS).[5]

The PDS maintains for current and future use a diversity of unique data sets acquired as national assets from across the Solar System after the expenditures of multiples of billions of dollars of

[2] As of 1 March 2017
[3] https://heasarc.gsfc.nasa.gov/docs/heasarc/other_archives.html
[4] https://earthdata.nasa.gov/about/daacs
[5] https://earthdata.nasa.gov/about

U.S. Government investment. This report summarizes an assessment of the current state of PDS, highlights community-based findings, and serves as a "roadmap" for its operation as it evolves in the next decades [Appendix A].

1.1 Establishment of the PDS

The NASA Planetary Data System (PDS) is the formal archive of the planetary sciences.[6] NASA established the PDS in 1989 to deal with concerns that the data being returned by scientific satellites was in danger of being lost. Problems included data storage media, adequacy of documentation, lack of data availability outside of the implementing science team, and the lack of any consistent standards for long-term data archiving. The National Academy of Sciences had chartered the Committee on Data Management and Computation (CODMAC) in 1982 [3]. Over the following six years, CODMAC issued three reports detailing means to address what were identified as serious problems in the way that NASA was managing its planetary data holdings [1, 7, 8]. Central among the recommendations was to have a scientifically guided distributed data system, adequately funded both to archive data and distribute it to researchers in a timely fashion. On the basis of peer-reviewed proposals, discipline-oriented nodes were selected to form the core of the PDS. These included [9]: Geosciences [10], Atmospheres, Small Bodies (asteroids, comets, and interplanetary dust) [11], Planetary Plasma Interactions [12], Rings [13], Imaging (focused on archiving large raw and derived imaging data sets and the ability to generate derived data) [14], Navigation and Ancillary Information Facility (NAIF) [15], and a Central Node for management.

The structure of the PDS has remained remarkably resilient through growth in the archive holdings and number of missions incorporated over the past 27 years. At the same time, technological changes in storage media, storage density, and accessibility of digital data have made tremendous strides, changes to which the PDS actively continues to adapt.

1.2 PDS Functionality

The Planetary Data System (PDS) archives electronic data products from NASA planetary missions, sponsored by NASA's Science Mission Directorate. The PDS actively manages the archive to maximize its usefulness, and it has become a basic resource for scientists around the world.

All PDS-curated products are peer reviewed, well documented, and available online to scientists and to the public without charge. Online search capabilities are also provided. The PDS uses standards for describing and storing data that are designed to enable future scientists who are unfamiliar with the original experiments to analyze the data. These standards address the data structure, description contents, media design, and a set of terms.

[6] https://pds.jpl.nasa.gov/about/pds_charter_12102015.pdf

Though the PDS does not fund the production of archive data from active missions, it works closely with project teams to help them design well-engineered products that can be released quickly.

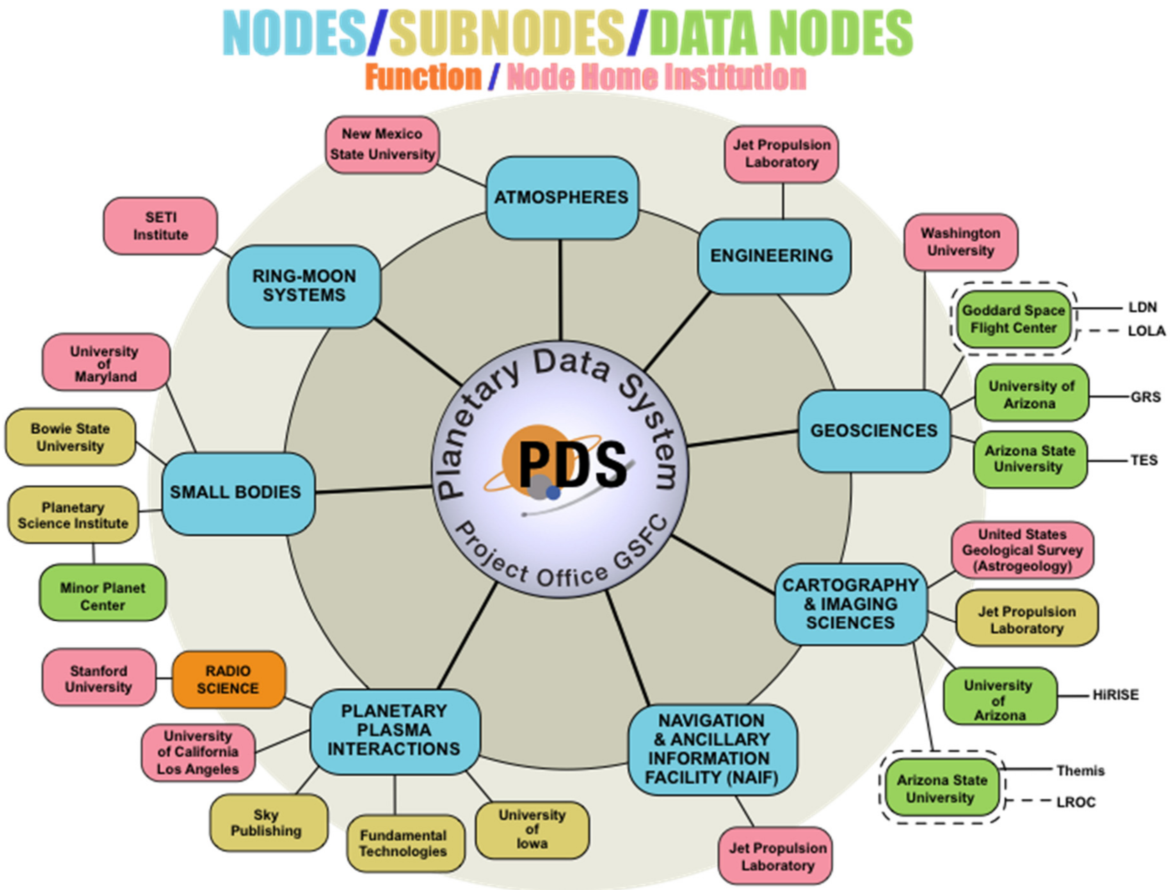


Figure 1. Current organizational structure of the Planetary Data System[7]

PDS Science Nodes now work closely with the community to archive higher-order data products (and other new archive materials) by supporting investigators in NASA's Planetary Data Archiving, Restoration, and Tools Program (PDART).

While PDS-curated products are freely available online, the PDS provides teams of scientists to help users select and understand the data. It also offers special processing for products tailored to the needs of individual users.

[7] https://pdsmgmt.gsfc.nasa.gov/nodes.html

1.3 Current Organizational Structure

The PDS continues to be organized as a federated data system [4]; data are archived by scientist-led organizations, called Discipline Nodes, which present a single interface to the world.[8] The organization of the PDS is shown in Figure 1. As in the original plan, additional functional groups provide engineering and user interface design services. Data Nodes are in use on an *ad hoc* basis to support various data intensive activities.

The current Discipline Nodes continue to be organized around broad areas—based on scientific discipline, as originally urged by CODMAC, by target body type, and by sensing modality. These broad areas reflect NASA's mission and the Agency's strategic plan for planetary science:

- Atmospheres (composition, structure, meteorology, and aeronomy) of planets
- Geosciences (geology, geophysics, surface properties, and tectonics) of planets
- Small bodies (comets, asteroids, dwarf planets, and also dust)
- Planetary Plasma Interactions (PPI) (solar wind-planetary interactions, magnetospheres, ionospheres, plasma tori) of planets
- Ring-Moon Systems (planets and rings and moons as dynamical systems)
- Cartography and Imaging Science (pushbroom imagers, hyperspectral imagers, analysis tools) of solar system objects

In addition, the PDS has two technical Support Nodes:

- The Engineering Node (systems engineering support, standards, technology investigations, coordination and development of system-wide software, and operations)
- NASA's Navigation and Ancillary Information Facility (It develops and maintains the observation geometry information system named "SPICE," used widely by both NASA mission scientists and engineers.)

The organization also has a Project Office that manages funding and budgets, and establishes common policies across the PDS.

1.4 Data Holdings

The PDS operates as a "living archive" of nearly one petabyte (PB) of planetary data managed by subject-matter experts and used by scientists around the world, and it is constantly expanding to allow new mission data to be made available to the planetary science community. With the complex data provided by increasingly sophisticated spacecraft and instrumentation, there exists a constant need to address the scientific priorities and challenges laid down in the U.S. Planetary Decadal Surveys, including new challenges of data preservation, data storage, and data access. Budgetary constraints present a constant challenge as well, making the situation more complex as PDS infrastructure grows to keep pace with rapidly increasing data volume.

[8] http://pds.nasa.gov

As PDS managers work to anticipate the technical and technological resource requirements for long-term data curation in planetary science, community guidance is needed to ensure that user needs continue to be met. Users of PDS data have come to expect increasingly more sophisticated data accessibility, and new user communities are developing as data management becomes a requirement across SMD programs. Constraints on data archives and services from International Traffic in Arms Regulations (ITAR) and data security issues will remain relevant and subject to unanticipated change. All of these challenges require PDS to maintain the ability to respond flexibly and rapidly to a changing data archive environment.

1.5 PDS Information Model

The most recent archive standard and information model for PDS, called PDS4, is in active development (Section 2.4) in part to address these challenges. PDS4 is designed to modernize the online data system and to improve the efficiency of data ingestion and delivery by supporting the scalable data storage and delivery architectures needed to provide access to ever growing data holdings. Because PDS managers and staff are actively engaged in the development and implementation of PDS4 in addition to their normal activities, not all of these PDS4 efficiencies are yet in place. PDS continues to manage voluminous PDS3 data holdings while supporting the delivery of numerous new data archives in both PDS3[9] and PDS4 formats. These circumstances reinforce the need for development of a "roadmap" such as this to guide PDS as it meets these challenges.

1.6 Establishment of this Roadmap Activity

During 2015, NASA re-competed and reviewed all PDS Node activities. The competition of all Science Nodes took place through the means of Cooperative Agreement Notice (CAN) NNH15ZDA006C (released 5 March 2015 and due 1 June 2015).[10] This activity was completed in September of 2015, and all of the selected Science Nodes are now funded for five years [Appendix H] with an option for an additional five years. The CAN process provides a governance model for PDS, similar to the "Institute Model" used elsewhere in NASA.

Performance Reviews were completed of the two internal nodes not competed through the PDS CAN, namely the Engineering Node and the Navigation and Ancillary Data Facility (NAIF), in January 2016.

Against this backdrop, a Planetary Data System Roadmap activity was established to look ahead to the time period from 2017 to 2026. The Roadmap activity began on 5 November 2015 with the release of a Request for Information (RFI) asking for community input by 5 January 2016.

---

[9] For missions whose Data Management Plans were negotiated before PDS4 became operational, the data continue to be archived in PDS3 format per those DMPs.

[10] https://nspires.nasaprs.com/external/viewrepositorydocument/cmdocumentid=449464/solicitationId=%7B6C202F60-D0E1-91E6-0D59-A6FC8C96EA36%7D/viewSolicitationDocument=1/PDSCAN Amend1.pdf

NASA chartered this current PDS Roadmap Study Team [Appendix B] to "develop a practical, community-developed pathway to implement the new long-term vision for the PDS, which continues to accomplish NASA's broad objective for the PDS; namely, preserving and making available all data products from planetary exploration research and missions."[11] PDS management produced the last Roadmap for the PDS [2006, Appendix C] and addressed the opportunities, technological venues, and associated challenges the system expected to face within the last decade. Self-assessments within PDS and a community survey have indicated that challenges identified in the 2006 Roadmap were addressed with varying degrees of success (Appendix D and Appendix E). A new Roadmap is needed to chart the future course of the PDS for the wide variety of PDS stakeholders (including data archivists, data providers from large NASA missions to small research investigations, data users, NASA managers, PDS staff, and the general public) for the next decade and beyond.

A Roadmap Study Team (RST) consisting of 16 individuals with differing backgrounds and interactions with the PDS was convened to consider what steps the PDS should take during the next decade to progress. Discussions began with the items in the RFI, namely, consideration of:

- Tools, resources, workflows, tutorials, and interfaces
- Making the archiving process seamless, less costly, and more efficient
- The role of PDS relative to other archiving alter-natives (e.g., journals), in providing the public access to NASA-generated data
- Integration of PDS data products and services with those of other facilities, e.g., the U.S. Geological Survey's cartography program and the Minor Planet Center, and other products
- The role the PDS should play in encouraging the development of higher-order data products
- Appropriate improvements to the current search capabilities of the PDS

The RST examined the Roadmap of 2006 in detail and interacted with the Nodes on that document, requesting and obtaining the self-assessments of progress against those plans during the previous decade.

This document presents the RST's understanding of the PDS as it now operates, as well as findings of fact, and suggestions for goals for the PDS and NASA to consider in furtherance of PDS activities during the upcoming decade.

2 Planetary Data System Background

2.1 What is PDS and What Does It Do?

The PDS is a distributed data archive that hosts and serves data collected by Solar System robotic missions and ground-based support data relevant to those missions. The PDS is managed by

[11] Dear Colleague Letter to Solicit Nominations for The NASA PDS Roadmap Team, 29 February 2016

the NASA Planetary Sciences Division as an active archive that makes available well documented, peer-reviewed data to the research community. "The main objective of the PDS is to maintain a planetary data archive that will withstand the test of time such that future generations of scientists can access, understand and use historical planetary data"[12]. The PDS tries to ensure compatibility across the archive by adhering to strict standards of data archiving formats and required documentation. The PDS is divided into science discipline "nodes" which are individually curated by planetary scientists [Appendix F]. The Solar System Exploration Data Services Office at the NASA Goddard Space Flight Center handles PDS Project Management.

The PDS is composed of eight nodes, of which six are science discipline nodes (Atmospheres, Geosciences, Cartography and Imaging Sciences, Planetary Plasma Interactions, Ring-Moon Systems, and Small Bodies) and two are support nodes (Engineering and the Navigation and Ancillary Information Facility (NAIF)). Engineering provides systems engineering support to the entire PDS, managing standards (data, software, documentation, operating procedures), technology investigations, coordination and development of system-wide software, PDS core-system and tools operations, leadership in relations with individual international space agency archives and the International Planetary Data Alliance. NAIF is responsible for implementing and delivering observation geometry data within a system called SPICE—a means for archiving, distributing, and accessing observation geometry and related ancillary data used in mission design, mission evaluation, observation planning, and science data analysis. In addition, there are several sub-nodes and mission data nodes whose exact status tends to change over time.

The PDS has close ties to the NASA Space Science Data Coordinated Archive (NSSDCA) that serves as the "deep archive" and long-term backup for validated PDS data holdings. The NSSDCA also hosts data products from earlier planetary missions and instruments not supported by the PDS and some non-NASA missions.

The PDS is a founding member of the International Planetary Data Alliance (IPDA[13]), a group supported by the international Committee on Space Research (COSPAR[14]), which actively works for common data standards and open planetary archives. The IPDA membership includes representatives from the space agencies of most spacefaring nations. The IPDA has adopted PDS4 as the international archiving standard for planetary mission data, and the archives of international missions are increasingly interoperable with PDS.

2.1.1 Comparison to EOSDIS

To understand how the PDS differs significantly from the far larger EOSDIS system, which supports data use and archiving in NASA's Earth Observing System (EOS), requires an introduction to the latter system. As noted in the Introduction above, EOSDIS relies upon 12 Distributed Active Archive Centers (DAACs), which are superficially analogous to the PDS nodes. However, EOSDIS is a more centralized organization, with stronger "top-down" management than

[12] See "PDS archiving philosophy" at https://en.wikipedia.org/wiki/Planetary_Data_System. From L. Feaga, PDS Small Bodies Node, University of Maryland 30 April 2008.
[13] https://planetarydata.org
[14] https://cosparhq.cnes.fr

is present in the "federated" PDS governance model. EOSDIS has established interoperability amongst the DAACs by adopting the Hierarchal Data Format (HDF), a system that has evolved to include airborne and campaign as well as satellite data. The centralized EOS Data and Operations System (EDOS) operates at NASA's Goddard Space Flight Center (GSFC). Processing of raw data is performed by six Science Investigator-led Processing Systems (SIPS). Both raw and processed data are transferred via a variety of NASA networks to the appropriate DAAC. A Common Metadata Repository (CMR) provides ingest and discovery capabilities for all EOS metadata. With all EOS data referred to the Earth, the smallest aggregation of data managed independently (called a "granule") is typically referenced to spatial and temporal extent of the data, but can also be tied to a particular data set from a particular instrument. As of March 2016, the CMR contained metadata for nearly 6,500 collections and over 270 mission science granules. The Unified Metadata Model (UMM)[15] was developed between 2012 and 2013 as a tailored information model, which melded legacy models with ISO 19115 at both the data collection and file level to support EOSDIS needs. The tailoring provided a significant gap in both philosophy and implementation from PDS4.

The fact that there is only one target—the Earth—along with well-known observational needs and research has enabled an efficient set of specialized tools to be developed. These include the Global Imagery Browse Services (GIBS), the Land, Atmosphere Near-real-time Capability for Earth Observing System (LANCE), and a Global Change Master Directory (GCMD). Significant infrastructure supports and enables data ingest via the Science Data Processing Segment (SDPS), configuration control, and metrics tracking via the COnfiguration Management EOSDIS Tool (COMET), and dedicated metrics tracking via the ESDIS Metrics System (EMS).[16] These systems are managed top-down by the Earth Science Data and Information System (ESDIS) Project at the Flight Projects Directorate of GSFC.[17] The far greater diversity of target bodies, instruments, and data sets precludes such a level of specialization within the PDS and also precludes a simple porting of PDS data holdings into the EOSDIS system, even if such a move were thought desirable.

However, the most significant difference between the implementation of the mission of the Earth Science Division and all other divisions within NASA's SMD is that EOSDIS not only manages all Earth Science data collected from all platforms (satellite, aircraft field measurements and others), but, more importantly—for the satellite segment—it provides for centralized command and control, scheduling, data capture and initial (Level 0) processing via the Earth Science Mission Operations (ESMO) Project for all instruments on all operating satellites, taken as a whole. By contrast, due to the diverse science goals and targets among platforms, the only comparable, centralized operational requirement on planetary missions is placed on deep-space missions, which communicate via the limited bandwidth asset embodied in the Deep Space Network (DSN) operated by JPL.

---

[15] https://earthdata.nasa.gov/about/science-system-description/eosdis-components/common-metadata-repository/unified-metadata-model-umm
[16] https://earthdata.nasa.gov/about/science-system-description/eosdis-components#ed-component-daacs
[17] https://earthdata.nasa.gov/about/esdis-project

The diversity of PDS data holdings, as compared with the EOSDIS holdings, can be illustrated by the observation that the 12,395 items referenced by the PDS search engine (on the PDS home page)[18] contain data from 98 different investigations using 44 instrument types with acquisition from 10 types of targets. In contrast, taking EOSDIS metrics from the Earth Science Division's annual reports,[19] FY 2016 shows 11,140 unique data sets with 3.21 million unique users and a total archive volume of 17.5 petabytes, growing at an average of 12.1 terabytes per day. There were 150.1 million files containing 2.937 petabytes of data ingested during the year (less the Ocean Biology DAACs for which data were not available). Thus, although the EOSDIS data volume is much higher than that of PDS (by a factor of ~18), it has only one target type (planet), and only one target (Earth). Conversely, EOSDIS contains data from 148 instruments (73 current and 75 historic) but spreads among only 9 instrument types, of which 8 are remote sensing and 1 (accelerometer) is *in situ*.[20]

2.1.2 Comparison to Other NASA SMD Archives

Other NASA Data Archives and Science Centers[21] have similar specialization implementations. The Space Physics Data Facility (SPDF) provides data from different instrument across 95 spacecraft and relies upon Common Data Format (CDF) software for data manipulation and display.[22] SPDF has over 1400 data sets with ~30 million files and > 25,000 parameters; total data holdings are ~120 terabytes, up from ~5 terabytes a decade ago [23] (numbers for the holdings of the Solar Data Analysis Center (SDAC) are not posted). The High Energy Astrophysics Science Archive Research Center (HEASARC), also at GSFC, coordinates data from high-energy astrophysics missions as well as data from the Legacy Archive for Microwave Background Data Analysis (LAMBDA). By 2010, HEARSARC holdings had passed 26 terabytes of data.[24]

Astronomical archive data, including that from the Hubble Space Telescope (HST) is collected at the Barbara A. Mikulski Archive for Space Telescopes (MAST).[25] This archive supports data sets from multiple instruments on 17 missions.[26] Data are typical stored as Flexible Image Transport System (FITS) files, which is under the control authority of the International Astronomical Union (IAU) FITS Working Group (IAU-FWG).[27] The most recent available MAST report[28] gives the data holdings as >500 terabytes, which is expected to grow rapidly once the James Webb Space Telescope (JWST) becomes operational.[29]

---

18 https://pds.nasa.gov
19 https://earthdata.nasa.gov/about/system-performance/eosdis-annual-metrics-reports
20 https://earthdata.nasa.gov/user-resources/remote-sensors
21 https://heasarc.gsfc.nasa.gov/docs/heasarc/other_archives.html
22 https://cdf.gsfc.nasa.gov
23 http://smd-prod.s3.amazonaws.com/science-blue/s3fs-public/atoms/files/2-7-SPDFToBigDataV2_2016Jun29a.pdf
24 https://heasarc.gsfc.nasa.gov/docs/heasarc/datadist.html
25 https://archive.stsci.edu
26 https://archive.stsci.edu/missions.html
27 https://fits.gsfc.nasa.gov/iaufwg/
28 https://archive.stsci.edu/reports/annual_report_2015.pdf
29 https://archive.stsci.edu/reports/BigDataSDTReport_Final.pdf

2.1.3 Context of PDS and NASA Within the Digital Universe

Roughly, the holdings across NASA's SMD can be estimated as:

- Earth Science Division 17.5 petabytes (17,500 terabytes)
- Astrophysics Division > 525 terabytes
- Planetary Science Division ~ 950 terabytes
- Heliophysics Division > 120 terabytes

No official "roll-up" is available for SMD, and this estimate leaves out some holdings for which numbers are not readily available, e.g., those for the various infrared data holdings in the Astrophysics Division, including those of the NASA/IPAC Infrared Science Archive (IRSA). Nonetheless, this does indicate that over 90% of the data holdings across SMD reside with the Earth Science Division. Even at the 20-petabyte level, such holdings remain small compared with those estimated for the worldwide "digital universe" of ~5 zettabytes (1 zettabyte[30] = 1,000 exabytes[31] = 1,000,000 petabytes). Of this, ~5% (that is ~250,000 petabytes) is estimated to have been analyzed and captured with appropriate metadata.[32]

There has been growing debate about how "big data" drives—and will continue to drive—storage, digital communications, analysis, and human-interface requirements [16]. Genomics may overtake astronomical science in the scientific realm as the driver in this regard [17]. Speculation has even turned to the potential future use of DNA as a storage medium [18, 19]. In any event, the relatively small size of all NASA SMD holdings (and the smaller size of PDS holdings within that) compared to the current size of the digital universe strongly suggests that PDS, while continually needing to be an early adopter and follower of best practices of IT in the coming decade, will not often be a driver of innovation needs or requirements.

2.2 PDS Requirements

PDS operates against a three-level set of requirements, which were re-articulated in 2005 and are periodically reviewed. These requirements (Appendix G) describe what the PDS must do and how it will be done. The Level-1 Requirements are:

1. PDS will provide expertise to guide and assist missions, programs, and individuals to organize and document digital data supporting NASA's goals in planetary science and solar system exploration.
2. PDS will collect suitably organized and well-documented data into archives that are peer reviewed and maintained by members of the scientific community.
3. PDS will make these data accessible to users seeking to achieve NASA's goals for exploration and science.
4. PDS will ensure the long-term preservation of the data and maintain their usability.

[30] https://en.wikipedia.org/wiki/Zettabyte
[31] https://en.wikipedia.org/wiki/Exabyte
[32] https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm

These requirements cover the essential activities of the PDS. They address the PDS goal of "collecting, archiving, and making accessible digital data produced by or relevant to NASA's planetary missions, research programs, and data analysis programs" as stated in the PDS Charter (Appendix I).

The PDS Level-1 Requirements are numbered, but we advance the interpretation that this numbering is for ease of reference and does not indicate any priority among the four Level-1 Requirements. Indeed, the numbering may have some correspondence to the sequence in which the requirements are met, as data must first be prepared, then validated, then distributed, and finally preserved for the long-term.

2.3 What Is PDS *Not* Intended to Do?

PDS serves complex, high-volume data from a wide variety of instruments to a large number of stakeholders spread across the globe. While PDS responsibilities are extensive, it is important to note that per NASA policy and practice there are numerous activities that are outside the purview of PDS. These include the following activities:

- PDS does not set the final archiving requirements for NASA missions. Mission program offices at NASA are the final authority for mission Data Archiving and Analysis Plans, including delivery schedules[33].
- PDS does not develop the data archiving or management plan requirements that appear in NASA Announcements of Opportunity or other announcements such as the Research Opportunities in Space and Earth Sciences (ROSES) or Cooperative Agreement Notices (CANs).
- PDS as an entity does not participate in the evaluation of proposed investigations.
- PDS does not develop pipelines for missions or perform pipeline operations for missions.
- PDS does not develop higher-order data products. Individuals at PDS nodes may propose to ROSES calls that include development of higher-order data products, but any funding for these activities is separate from PDS funding.
- PDS does not actively participate in non-U.S. mission data archiving unless or until there is a NASA MOU or other agreement.
- PDS does not accept ITAR-restricted data.
- PDS does not archive software (see Finding X).
- PDS does not curate geologic samples, analog materials, or physical objects.

2.4 Assessment of Progress Relative to the 2006 PDS Roadmap

In February 2006, PDS Management finalized a presentation entitled "PDS Strategic Roadmap, 2006–2016." The slides from this presentation are included as Appendix C in this report. Slides 19 through 22 of this presentation set out a series of "five-year goals." This previous Roadmap is a useful benchmark to evaluate the progress of PDS over the past decade. In September 2016,

[33] Although PDS does set the requirement for PDS4 use and details of archive delivery, NASA sets the archiving requirements *per se* via AOs, NRAs, CANs, and associated contracts and grants.

the PDS Discipline Nodes were asked to comment on their views of the progress that had been made in light of the 2006 Roadmap goals. Appendix D summarizes the responses.

The 2006 Roadmap also included a vision for PDS4 (Slide 23 in Appendix C). This major re-imagining of the PDS infrastructure would include, among other things, "a simplified set of data formats," "metadata requirements to support modern search technology," "highly automated validation and ingestion," and "improved interoperability with other domestic and international space data systems." Implementation of PDS4 has been a major accomplishment of the past decade for the PDS. Although the final implementation is still in progress, it is already apparent to the nodes and to new data providers that the PDS4 standards and systems represent a significant improvement and substantial investment in the future of the PDS archives.
Current PDS Structure and Implementation Plan for the Next Five Years
The current PDS Science Discipline Nodes were chosen competitively in 2015 through a Cooperative Agreement Notice (CAN, NNH15ZDA006C). Each proposing team was required to identify PDS mission data sets from past, present, and future missions within the total PDS archive that the team proposed to archive and to curate, including restorations. The PDS CAN included a list of all past and current mission data sets as well as data sets from those missions in phase B or beyond in 2014. The accepted proposals led to the same ensemble of discipline nodes that has comprised PDS since 1990. The selected member nodes, together with EN and NAIF, were used to administratively reconstitute the PDS structure, implement their operations plans (Appendix H), and address the new PDS charter (Appendix I) in September 2015.

3 The Challenges Facing PDS Today

3.1 User Needs and Expectations

The PDS user base and its expectations have changed dramatically over the last decade. In fulfilling its archiving role, the PDS was primarily designed to provide storage and curation of data over time, but part of its charge was also to allow contemporary scientists to access its holdings in order to enable research work. The last decade has been an amazing time for Planetary Science, with a tremendous number of missions that have returned data, and with a correspondingly increased number of scientists that want to access that data. That success, combined with the rapidly expanding capabilities of computers and computer networks, has also empowered the non-science general public to be interested in those very same data.

The population of PDS users has expanded to include more professional scientists than ever before, encompassing a wide variety of experience and expertise with planetary data, as well as members of the general public seeking access for an ever wider potential array of uses.

In the last decade, the way that everyone interacts with digital information has undergone an evolution as well. Widespread internet usage has conditioned users of all varieties to expect more than just a static storage archive. They seek interactive services that will help them explore, discover, sort, and visualize data before they even download the data locally.

Part of the challenge for PDS, and for this Roadmap, is to consider how to respond to these changes. A significant increase in institutional support, including but not limited to increased funding, may be required if PDS is to meet the steadily rising reasonable expectations of its user community. If a decision is taken not to provide the increased support PDS would need to meet certain expectations, then such a decision and its rationale should be clearly stated by NASA. Conversely, some of the heightened expectations may be unreasonable, in which case it is imperative that all stakeholders clearly understand the situation.

3.2 Data Discoverability

Users typically expect to be able to search for data based on what they already know: data source (mission, instrument); characteristics (e.g., image, wavelength); target name (e.g., "Mars") or type (e.g., "asteroid"); or a specific reference (e.g., DOI, URL) found in a publication. This is a reasonable expectation, but satisfying it depends on the quality and uniformity of metadata in product labels, and also on the search services developed by PDS and the nodes individually. The PDS4 system does provide a much greater degree of control and assurance over these search terms in product labels, as well an enabling the expansion of available search terms for more specialized use in subsets of the data.

It is often stated that users want to be able to search across the entire PDS archive with one query, but in fact the bulk of actual end-user queries do not cross even one node boundary, let alone all six, just by the nature of the discipline-specific distribution of the archive. Still, it would not be unusual for a single query to involve two or perhaps three nodes that all have a discipline interest in a particular data collection. So, for example, a user investigating rings around asteroids might run his query at the Ring-Moon Systems Node, but would also want to search the asteroid holdings at the Small Bodies Node simultaneously.

One concern is that a user should not have to know where the data are archived in order to discover them. A general search interface that does reference all node collections is maintained at the Engineering Node, as part of the PDS system services, specifically to handle these broad queries and direct users to the relevant node interface where the user can perform various orders of queries on the specific collection of data products at the reference node.

Users want some form of quick-look or visualization support to confirm their interest in a data product before downloading it. The era of downloading 10,000 files in order to get to 10 files of interest is over. Users, in general, want to download only what is actually relevant up until the point where it is faster to download everything than try to discriminate (a fuzzy boundary, at best). But determining relevancy from search results can be difficult if the only available information is a brief description from a label, especially if every one of those 10,000 labels has essentially the same description. Quick-look visualizations, iconography, "Read more…" options, and the ability to sort and filter results based on summary properties ("facet-based searching") all assist users in getting to the relevant results, and they are now so common generally that users expect similar support from sites like PDS.

3.3 Data Usability

Users want to be able to find documentation that is specific to the data they selected (filter curves, mounting diagrams, metadata definitions, etc.). Building on the previous item, supporting documentation has always been important, but as remote instruments become more sophisticated and their data more complex to analyze and interpret, connecting data products to the ancillary information and descriptions needed to fully understand the observation in hand is critical. Users will need to go back into the archive to retrieve additional documentation based on references in downloaded data. Therefore, it is critical that documentation in the PDS archives be as locatable and accessible as the data products themselves.

It should be noted that the amount of documentation supplied with data by missions and other large data providers has been increasing over the years and now represents a substantial part of the archiving effort and a major contribution to the usability of the archive. Documents ranging from planning data and observing logs to detailed calibration reports, user guides, and even introductory documents prepared by science team members are all being included in the archive and present a boon to users now and in the future.

Users, however, do not want to have to read substantial documentation before being able to locate data. That is, users want the search services to be intuitive at least for the most common types of searches. Given the state of search services in the industry, this is entirely reasonable. This is the capability currently provided by the data search option on the PDS home page. Users type text into a search engine box, and the underlying service compares terms to words indexed from all metadata fields to generate a return set.

Users want to receive help using PDS data and tools. This includes users who want immediate help to overcome a very specific problem, users who want or need a deeper understanding of a particular data product or collection, and users who want to learn more about the PDS system and archives as a whole. The solution needs to address many levels, from email addresses and phone numbers on contact pages of a website to introductory documents included with data submissions to tutorials and workshops.

Conversely, many users want to find simple but adequate instructions without needing to communicate with PDS personnel.

3.4 Tools and File Formats

Analytical tools evolve quickly. New file formats typically arise after a period of instability and competition among solutions. Existing formats can fall out of common use over just a few short years when something better comes along. To attempt to stay abreast of format evolution even within a single discipline would be a daunting task for PDS as an archive; to do so for all of the planetary science disciplines that PDS serves is not feasible. This is why PDS4 formats have been designed to be robust over the long term and difficult to *mis*understand, even decades from now when computing standards will have dramatically changed.

The challenge to PDS is that data providers, data users, and the PDS itself all have competing priorities when it comes to data file formats.

Data providers prefer to submit the data in the format in which they produced it. This is partly an economic consideration—it takes time and space to code a format translation routine and verify that it is properly transforming without corrupting or degrading the data. Also, since the provider's format is typically used directly by the science team in generating published results, the team has a high degree of confidence in the data provider's format.

Data users do not want to have to reformat data themselves before they can use it. The processing and analysis tools that users want to utilize have defined input formats, so a desire to see PDS provide data in those formats is understandable. Format translation can be a tricky proposition. If PDS were to perform the translation, users would have a high degree of confidence in the result, perhaps more so than if they ran a translator themselves.

PDS wants archival formats that are simple to support, without requiring format migration to preserve usability. Performing format migration of archived data is necessarily a slow, careful, and expensive process. Even the process of transforming data between contemporary formats must be done with some care to avoid inadvertent corruption or degradation. PDS provides data to users in the archive format or in one of a small set of supported transformations. PDS personnel have expertise in the data object and metadata structures of the archival form, which is the format that works with the tools PDS provides.

3.5 Online Processing and Analysis

Users want to be able to send PDS data directly to online analysis tools without having to download and upload. PDS services like the Cartography and Imaging Sciences Node's Projection on the Web (POW) provides users with the ability to identify planetary image data using the Planetary Image Locator Tool (PILOT) and route it to POW for further processing. This kind of capability is part of the evolution of processing. Also, while it constitutes a small part of planetary activity now, projects like VESPA (Virtual European Solar and Planetary Access) have the goal of providing software interface layers on archives like the PDS that allow users to select files and send them directly to web-based analysis tools for processing, avoiding the separate steps of downloading, uploading, and possibly running a format conversion in between. The highly constrained PDS archival formats and well-defined metadata present a very attractive basis for this sort of interface.

3.6 Increases in Data Volume, Variety, Complexity, and Number of Data Providers

A mission's archive data volume depends on the particular mission objectives and instrumentation. It is certain that as missions grow ever more sophisticated and capable, and new instrumentation is built to meet mission objectives, the volume of returned data (raw, reduced and calibrated products) will increase. In addition, as NASA data analysis programs increasingly require archiving with PDS, archive data volume within the PDS will grow.

Data volume growth is only one consideration in planning the archive life cycle. As the number of data providers will increase in coming years, each will require some level of training in data preparation and archiving. Finally, the increasing variety and complexity of the data to be archived will result in new metadata unique to particular instruments and processes that will need to be captured. In many ways, the effort required of PDS node staff scales with data diversity more than with data volume.

The PDS4 standard and system were developed to address the impact of each of these expected growth areas. The PDS4 registry supports discoverability across the growing volume of archived data. PDS4 data dictionaries provide the flexibility required to accurately capture new metadata descriptors. PDS4 documentation and tools will assist data providers in creating, testing, and delivering data to the archive.

3.7 Laboratory Data and Physical Samples

The Curation and Analysis Planning Team for Extraterrestrial Materials (CAPTEM) responded to the PDS RFI (NNH 15ZDA012L), as summarized in Appendix J, and met with Dr. Tom Morgan to discuss their response. Laboratory analysis of samples returned from NASA missions depends upon the stewardship of all mission-returned samples by the Astromaterials Acquisition and Curation Office at Johnson Space Center (JSC) curation. These include Apollo Samples (first returned in 1969), Genesis Solar Wind Atoms (2004), Stardust Comet Particles (2006), Stardust Interstellar Particles (2006), Hayabusa Asteroid Particles (2010), and will include Hayabusa 2 Asteroid Particles (2020) and OSIRIS-REx Asteroid Particles (2023).

Decades of data have been collected on these samples in laboratories with increasingly sophisticated and efficient instruments. Little or none of the scientific data obtained on any of these extraterrestrial samples is currently, systematically archived in a "PDS-equivalent archive" as defined in program element C.7 (PDART) of ROSES 2017. Older, non-digital data exists in publications and preserved notebooks. More recent and current data are often stored in online supplements to publications, voluntary archives,[34] and individual principal investigator (PI) digital storage. Indeed, many of the early PIs of the Apollo samples are retired or deceased. Collecting, validating, and archiving these data would be a large job, and significant resources would be required to ensure its success.

Also curated by JSC curation are astromaterials samples that are not mission sourced, including Antarctic Meteorites (first collected in 1976, with the Smithsonian Institution's National Museum of Natural History), Cosmic Dust Particles (first collected in 1981), and the Micro-Particle Impact Collection (MIC, formerly called Space Exposed Hardware, first collected in 1985). Additional samples important to planetary science include meteorites curated at private museums such as the Field Museum (FMNH) and the American Museum of Natural History (AMNH). Future quasi-private or fully private space missions may return high-value samples yielding important planetary science data.

[34] e.g., https://wiki.duraspace.org/display/DSPACE/Home

While spacecraft mission teams must meet PDS delivery requirements, include funded data-professionals, and produce relatively uniform products of planned size, most sample analysis is performed by individual PIs. A moving target in discussion of extraterrestrial samples is the requirement of a data management plan (DMP) for PIs proposing to Research Opportunities in Earth and Space Science (ROSES) programs. A systematic plan is not yet in place that would address the DMP needs of ROSES PIs working on extraterrestrial samples. However, many ROSES PIs working on samples presently lack the incentive to archive data properly and do not possess the data management resources present in mission teams.

3.8 PDS Structure and Governance

As discussed in Chapter 2, the beginnings of the implementation of the PDS trace back to a NASA NRA of 1989. Over the past 27 years, there has been evolution of the Discipline Node structure, funding vehicles, and management approaches. As technical needs and implementation approaches evolve, so do management needs to support the overall requirements on the organization. At the very least, periodic review of the PDS management structure both by (external) NASA management and (internal) PDS management, are a good idea for ensuring the continued efficient running of the PDS while effectively serving the planetary science community and other current and potential stakeholders.

4 Issues, Observations, Findings, and Suggested Remediations or Actions

In the context of the history, background, and previous decade performance of the PDS, the Roadmap Study Team identified 13 broad issues and observations for which there are an associated 18 “Findings.” Of these, two can best be described as “stay the course,” while for the rest the RST identified potential actions and/or remediations, which could be taken up directly by appropriate managers and/or considered in more depth by others, e.g., the Planetary Science Advisory Committee, or PAC, as deemed appropriate by NASA Planetary Science Division management.

These observations and statements of issues follow, roughly in order of importance as per Chapter 3.

4.1 User Needs and Expectations

Finding I: PDS Stakeholders

In the overview section of the PDS CAN under which the current PDS nodes were selected and funded, there are repeated references to "researchers" and "the science community", but there are also references to "the public" as potential benefactors of the PDS.

While it seems reasonably clear that the primary stakeholder in the PDS core activities is the scientific community, that community is itself divided. For example, node objectives in the PDS

CAN include both advising missions preparing archival data and supporting researchers using the archived data (the "contemporary science community") and curation of the data for use by successive generations (the "future science community").

As the connection point among various communities, PDS can be caught in the middle of competing and potentially conflicting interests.

Finding I: While all PDS stakeholders are recognized as valuable, the prioritization of stakeholder interests and the impact those interests should have on PDS policy, design, and resource allocation are unclear.

Remediation in this case might include a clear statement from NASA indicating the various stakeholders in the Planetary Data System, with a prioritization that can be referenced to provide guidance when conflicts arise.

Finding II: Managing Expectations of PDS Usability

There has been a history of complaints both by data consumers and data providers of inadequate capabilities of PDS. In many cases, these complaints are rooted in evolving user needs and expectations over the nearly three decades since PDS was chartered (Section 3.1). While user expectations for interactions with the PDS vary, the fact that there is a mismatch between user expectations and PDS capabilities is a problem.

The 2006 Roadmap set high expectations for what it hoped PDS would accomplish during the ensuing decade. Many of these expectations have been met, but others have not.

These factors illustrate an ongoing fundamental mismatch between some desired goals of users and the resources available to achieve them. Through informal internal perspectives from PDS Node personnel, and interaction with PDS overall, the Roadmap Study Team has concluded that the PDS's current funding profile does not support the very high expectations of users now and may be unlikely to do so in the future. Management faces a fundamental choice: either allocate resources and funding appropriate to achieve the desired goals, or significantly scale back the stated goals to fit the current funding profile, either by ignoring new goals or giving short shrift to old ones. Typically, the PDS has maintained, and even sometimes elevated, the stated goals for PDS over the years, even when adequate funding was not provided to achieve those goals.

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 NASA management.

A suggested remediation would be that NASA either allocate additional resources to PDS, to bring PDS services more in line with user expectations, or that NASA work with PDS to educate the public about the functions that the PDS can accomplish under the current resource guidelines.

4.2 Data Discoverability

Finding III: Data Discoverability

When PDS was first designed, a user who wanted to look at images of Enceladus would arrange for PDS to send them the entire Voyager imaging data set on a series of CD-ROMs, and then the user would take it upon themselves to browse through the data set and find the images they wanted. Today, the PDS holdings are far larger and more varied. Furthermore, users are far more accustomed to finding what they want with only a few clicks of a mouse.

Several of the PDS Nodes have online search tools that are very successful at serving their user communities. However, less experienced users may not know where to find these tools, and they might not even know which Node they should visit in order to find the data they want (Section 3.2).

Finding III: 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.

There are many potential avenues that could be pursued in order to address different aspects of this finding.

The PDS4 metadata provides substantial support for searching at various levels, and the PDS4 registry system provides a new set of tools and techniques that can be employed in developing search capabilities (see Finding VI). Quickly directing users to the service most likely to meet their needs would be a challenge and likely require a deeper understanding of how various types of users want to find and use PDS archive data. Additional cues to help users filter and select (or reject) results, such as thumbnail images, are becoming increasingly important as the size and breadth of the archive holdings continue to increase.

There are other ways to refine the PDS user experience. The PDS could conduct focus group sessions with end users who are new to PDS and observe them as they attempt to find data products in PDS. This has been done in the past, but state-of-the-art web design requires periodic refreshment, and end-user testing of the revised web design would help to confirm that it successfully addresses the problems identified by the initial tests.

Another approach is that emerging search technologies may provide new ways to evaluate the potential relevance of query results. PDS can then work to improve its web presence and search capabilities in such a way that users are more likely to find what they seek. The PDS could have and make use of input from a PDS User Group from across the community.

Additionally, improvements to documentation and training (as mentioned in Finding VIII) could improve the user experience.

Finding IV: Integration with Other Archives

The PDS occupies a national and global leadership position in setting standards for archiving and serving planetary science data. PDS constraints on archival formats are well justified by the mission of the PDS and its mandate from NASA. Where necessary, PDS nodes make reasonable compromises to accommodate both the analytical needs of data pipelines and archive format constraints.

The PDS has developed interfaces with NSSDCA, PSA, and other agencies. Searching the PDS archives for Rosetta data, for example, will return results from both the PSA and the PDS archives. The successes of IPDA and other efforts have demonstrated that PDS4 is an interoperable standard. Future opportunities include interoperability with the sample catalogs at curation facilities, the archives of long-term missions like the Hubble or James Webb Space Telescopes, and various international archives.

Interoperability is a paramount long-term goal. The potential is a planetary archive ecosystem that allows users to easily find all the data sets relevant for their research. For example, MESSENGER data in PDS would be searchable in the same query that also searches the (future) ESA mission BepiColombo data. A single query on “Fra Mauro formation” would not require searching PDS, JSC curation, Russian curation, etc. The PDS4 service structure and APIs provide a strong foundation for spreading the same sort of interoperability interface across the global planetary science community.

Finding IV: The PDS serves as the model for other national space-mission data archives in ensuring future universal accessibility and searchability. The PDS is uniquely poised to lead efforts to make national and global archives interoperable.

No remediation necessary. PDS has found significant success in this regard and should keep moving forward.

Finding V: Citation of Data Sets

There is a growing desire in the planetary science community and among its journals to be able to treat data sets on a par with refereed journal articles in terms of citation and referencing. The effort to produce a peer-reviewed, archived data set like those curated by PDS is substantial. The typical archival data set is well beyond the size that can even be contemplated for actual publication in a journal or supplement. Notwithstanding, the principles of accountability and reproducibility demand that when a result is based on a data set, that set be both clearly identified and available for public inspection and re-analysis. Furthermore, citation and reference counts for refereed publications (article or data set) provide metrics on usage and impact that are verifiable, and in the case of data sets can be critical to authors who have made archiving a significant part of their scientific career activities.

Recently, PDS undertook a pilot program to analyze how its archive holdings have been and might be cited in the literature, and to develop a process for obtaining permanent digital object identifiers (DOIs) for PDS products. In conjunction with the U.S. Department of Energy Office of Scientific and Technical Information (OSTI) and DataCite (a DOI service provider focused on data citation), the PDS will be acquiring DOIs (through OSTI) and setting up a local service based on the PDS4 registry to resolve DOI queries for PDS products. The metadata design for PDS4 labels already aligns very well with the metadata required and recognized by DataCite, making the prospects for automation of the DOI registration process very bright.

In addition, PDS nodes involved in the pilot program have been in contact with the journals manager of the American Astronomical Society and the PI and IT specialists at the Astrophysics Data System (ADS), to coordinate with them the mechanics of including data references in papers, and incorporating the data product DOIs and associated metadata into the ADS database in order to generate reference and citation metrics for PDS data.

The final piece needed to complete the connection between data sets and resulting publications will be education and publicity: authors need to be encouraged and reminded to cite the data they use; referees need to develop the habit of demanding to see data references included in analytical papers; and journal editors need to make it a matter of policy that complete data set references must be included in the articles they accept and publish.

Finding V: PDS is actively involved in addressing the data citation issue, and is well-positioned to provide the essential links in the chain needed to enable clear, direct referencing of PDS products; but it cannot itself change the habits and attitudes of authors, referees, and journal editors when it comes to including data set references in publications.

Remediation for this needs to come from several sources. However, it might substantially help the process if NASA planetary data analysis programs clearly stated that publications resulting from analysis of existing data sets should clearly and formally reference those data sets, and if the guidelines for proposals submitted to NASA to analyze existing data sets emphasized the need to formally reference those data sets in the proposal itself—perhaps providing examples until data referencing becomes commonplace.

4.3 Data Usability

Finding VI: Modernizing Metadata

The PDS has undergone considerable evolution since its initial establishment in 1989 based upon the recommendations of the National Academy of Sciences via the Committee on Data Management and Computation (CODMAC), chartered in 1982. Data services, standards, and hardware have evolved significantly since then. Gradually the standards and approaches to metadata for archiving have matured, and the PDS has successfully endeavored to keep pace with the changing standards. During the last decade, the updated PDS4 data model has been brought into use, the advantage of which is its strict adherence to and compliance with modern archival international standards [20]. However, it should be noted that for the immediate future the bulk of the data in the PDS, both legacy data and data coming in from mission pipelines developed prior to the introduction of the PDS4 standards, is in the PDS3 format.

Finding VI: The accessibility and discoverability through the PDS4 metadata registry is a cornerstone to the future of community interaction with the PDS as a coherent storehouse of data. Legacy data archived in PDS3 format (the vast majority of PDS holdings) often lack metadata sufficient to enable discovery and accessibility commensurate with PDS4.

A potential remediation for this is conversion of PDS3 data into PDS4 format, or at least the generation of PDS4-compliant metadata for PDS3 data sets. This has already been done for some data sets (e.g., Voyager ISS), but it is not a trivial task and would require substantial effort for data from many completed missions, as all the metadata needed for PDS4 may not be available. Accomplishing such a task would require a dedicated allocation of resources, whether directly to PDS or through other funding mechanisms via ROSES, like PDART.

Finding VII: Access to Data

Most PDS data are available online and can be downloaded.[35] However, while the submitted data archive files are available, more could be done to facilitate discovery and analysis of the data.

The PDS Engineering Node has provided an API to the registry and search services to allow third party development of queries and services and also to support communication between PDS and other OAIS/ISO-compliant archives. Such an approach promises to eventually provide convenient access to PDS holdings. The PDS4 uniform metadata standard[36] enables such increased data service. The intent is that all metadata submitted to the PDS can be ingested into PDS registry database(s), which would facilitate searching.

Finding VII: The PDS does an excellent job of providing access to its data holdings and is on track to increase such access. The latter is enabled by the PDS4 uniform metadata standard.

The Roadmap Study Team encourages the continued development of the registry and its API. Such an implementation could, for example, serve as the foundation for more complex or higher-order data services, for desktop or mobile users.

Finding VIII: Documentation and Training

The number and types of users PDS supports—and is expected to support by the community—has grown and evolved and continues to do so. The Roadmap Study Team expects that this will

[35] Some older datasets are held by PDS but not fully archived. These "safed" data sets are available by request. For more information, see http://pds-imaging.jpl.nasa.gov/holdings
[36] https://pds.nasa.gov/pds4/doc/sr/current

continue and not slow during the near future. Users and their predominant questions include, but are not limited to, new data providers ("How to do it?"), casual programmers ("What does it look like?"), and reviewers and end-users ("What do you have, how do I find it, and how do I interpret/use it?"). There are also experienced end users, such as pipeline programmers looking to incorporate PDS software directly into their pipelines.

In addition to the PDS4 Standards Reference and Information Model documentation already available from PDS, documents to address special topics of concern to data preparers and users would be useful (Section 3.3). Such topics range from how to estimate costs of producing a mission archive to how to read the label of a PDS product a user has downloaded from the archive. Given the range of potential PDS users, "documentation" may include tutorials, worked-example sets, or even video presentations on specific topics, in addition to the more traditional, formal document. Published and peer-reviewed documentation ensures the PDS has common expectations of data providers, reviewers and users across all nodes. Providers, reviewers and users have varied, node–dependent experiences with data submission, review and support due to lack of documentation and agreed upon "best practices."

Good training is essential to the long-term success of the archive. Training opportunities range from on-site training for individuals or small groups, to sessions at meetings where potential data users and providers are in attendance, to video recordings of training sessions, which can be viewed online.

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.

A suggested remediation is for NASA to encourage and provide resources for the development of user documentation, tutorial materials, and hands-on workshops when cost-effective opportunities arise. New funding for an annual U.S. counterpart to the recent PDS4 workshop in Spain[37] could be very effective.

Regular opportunities for training in data access, analysis, visualization capabilities are provided by the Planetary Data Workshops[38] [21], as well as at a wide variety of independent mission-specific training events at national meetings such as the AGU and LPSC.

4.4 Tools and file Formats

Finding IX: PDS File Formats and Translation Software

The PDS archival file formats are simple to support across generations (human and technological) without requiring format migration to preserve usability. Users, however, have different goals in mind for the data once they obtain it from the archive, and the archival formats are often not easy

[37] https://www.cosmos.esa.int/web/psa_pds4_workshop
[38] http://astrogeology.usgs.gov/groups/planetary-data-workshop

to work with in their analysis tool of choice (Section 3.4). Given the work of PDS as an archive, limited resources are available within PDS to support format transformation, even though doing so supports its Level 1 requirements to make its data accessible and usable. So, as a practical matter, PDS could focus on transformations that are broadly applicable to large sections of the end-user community and likely desirable for a relatively long period of time.

There are standards—like JPEG, GIF, IAU-FITS, or VOTable—that are defined by a recognized standards body, have a lifetime that has or is likely to be measured in decades, and provide general compatibility with a large number of processing environments. For these standards, it is reasonable for PDS to invest in developing specific tools to reformat archive data where appropriate. Users *must* be informed in all cases where format translation might result in a loss of data fidelity (e.g., transforming from a 16-bit PDS image to an 8-bit GIF image).

One of the major goals of the PDS4 redesign, and in particular the metadata design, was to make it feasible for third parties to write software that can handle large sections of the archive with confidence, without needing intimate *a priori* knowledge of the data in the archive and without having to update code every time a new data-provider submits data. The PDS Information Model is the essential foundation for supporting third-party development.

PDS is expert in its own standards. While it is not an end-user application development house, it certainly has the expertise to advise developers on its format standard; and because active scientists manage the discipline nodes, it often also has personal contacts with members of the programming community. These connections can be utilized to address the interface needs of both data providers who use the tools to produce archival data products, and end users who want to use archive products in their analysis. It is critical that the programming community sees the PDS4 format as accessible, reliable, and amenable to coding solutions.

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.

A suggested action to address this finding is for the PDS to support format translation from PDS4 format to a few selected file format standards, and also to provide encouragement and support to non-PDS software developers to include PDS4 input/output capabilities in their tools. PDS has made a good beginning to this task but more work remains.

Finding X: Archiving Software

The data archiving mandate from NASA, which lies at the foundation of PDS, is not merely to preserve bytes, but also to preserve the usability of the data for future generations. Consequently, PDS has developed data format standards, metadata standards and requirements, and rigorous review and validation procedures for archive submission and acceptance. The PDS data standards ensure that archived data are independent of both software and hardware environments. All of the organizational expertise, experience, standards, and support structures of PDS are specifically geared toward supporting and serving its archival data products.

More recently the question has arisen of possibly using PDS to archive software.[39] To archive software to the same level of support and curation as PDS does for data would require expertise in code analysis, external peer review standards for code, the development of specific requirements for coding and both programmer and user documentation, and the maintenance of expertise in-house at PDS to aid and advise code users. PDS has none of the necessary organizational expertise, personnel, or machinery to support such an endeavor.

To act as a repository for software, PDS would have to, at the very least, provide the same level of support that a repository like GitHub does—requiring and posting contact information for external coders who are expected to provide support for at least some period of time, monitoring and protecting the repository against malicious changes or code loss, providing support to search through the code libraries, documenting versions and package dependencies, and so forth. PDS has neither the expertise nor the infrastructure to support this sort of activity.

As mentioned previously, there are already issues with user expectations for PDS versus the actual services provided by PDS with respect to archival data. Users who sought software in the PDS would reasonably have an expectation of some level of support and might also expect that software in the PDS would be supported at the same level as data in the PDS, irrespective of labels like "archive" versus "repository." The fact that PDS is not organized, funded, or staffed to meet those expectations would only add to the population of dissatisfied PDS users.

Finding X: The PDS is not an appropriate archive or repository for software.

Remediation in this case could start with systematically removing any text implying that software should be either "archived" or "deposited" with PDS from all PSD AOs and RFPs. Because the subject of how to preserve and build on the investments being made in analytical software both inside and outside of mission pipelines seems to be arising more frequently and in more contexts, additional remediation might also include the instigation of a study to consider the question of what might constitute "software archiving" and if such a thing might now be feasible within current technology. This might be considered as a similar effort to the Planetary Data Workshops organized in the early 1980s that led the eventual creation of the PDS itself.

4.5 Online Processing and Analysis

Finding XI: Information Technology

The PDS overall, as well as the individual nodes, continues to be proactive in examining information technology that would help to maintain lower costs while improving efficiency for users

[39] For example, ROSES 2017, in Section 3.6.1 of Appendix C.1, while giving specific guidance on how certain software created as part of a NASA award might be made publicly available, says, “NASA expects that the source code, with associated documentation sufficient to enable the code’s use, will be made publicly available via GitHub, the PDS (for mission-specific code, when appropriate), or an appropriate community recognized depository (for instance, the homepage of the code base for which a module was developed). Archiving software in a public repository does not require the proposer to maintain the code.”

and providers. An example that continues to be considered is the use of "the cloud", but current PDS node estimates show this not to be cost effective for PDS usage, mostly due to current data-access fee structures. The PDS is well prepared to move into cloud computing and cloud storage through the PDS4 architecture if and when this becomes cost effective. PDS data holdings and usage requirements are small compared to many other users in the "Digital Universe" (section 2.1.3). However, driven by the diversity of PDS holdings (cf., e.g. Section 2.1.1) and OAIS compliance enforced via the PDS4 uniform metadata standard, the PDS has been a leader and innovator in the field of metadata definitions and open archive interoperability [22–27].

Finding XI: The PDS has been and continues to be proactive in investigating information technology and adopting best practices.

No remediation needed.

4.6 Increases in Data Volume, Variety, Complexity, and Number of Data Providers

Finding XII: Potential Impact of ROSES Archiving Requirements

The existing PDS structure was designed to ingest large amounts of data from spacecraft missions, which could provide their own dedicated data scientists to work on creating archives to be submitted to the PDS. There is a trend towards increased archiving requirements for individual researchers funded by ROSES programs[40], though at this time PDS archiving is not required for most ROSES programs. The new ROSES requirement to archive data potentially creates a new category of submission and submitters to the PDS. Typically, these individuals would not have archiving experience and may need additional training by node personnel in order to have their submissions accepted into the PDS. This would represent a new, significant, underfunded burden to node staff.

The critical need for training precedes the point of archive preparation. Potential non-mission data providers, such as those proposing to ROSES programs, as well as small-satellite missions, balloon and airborne observations, and ground-based observations of planetary targets, would be in a much better position to formulate Data Management Plans (DMPs) if they had access to some basic training in data preparation planning and procedures. The preparation of DMPs is evolving as NASA program requirements change, and PDS is supporting these potential data providers with online information and training materials[41] that are updated annually. This activity is crucial, but currently insufficient for proposers to meet ROSES requirements.

Finding XII: It is a matter of concern as to whether the PDS nodes will have the resources to serve the data archiving requirements of individual ROSES investigations.

[40] ROSES 2017. Planetary Sciences Program Overview. Section 3.6. C.1-6.
[41] e.g., http://pds-imaging.jpl.nasa.gov/help/proposals.html

Suggested remediation is for NASA to ensure PDS has sufficient resources to support non-mission data providers in a timely manner. These additional resources could be used both to create more tools that would ease data submission and validation, and also to employ more PDS archivists that can create training materials and assist ROSES investigators in their submissions.

Finding XIII: Higher-Order Data Products

Many mission teams produce higher-order data products as part of the process of understanding and analyzing their mission data. Such items include derived products such as maps, co-added images, and products to which additional calibration or data-reduction processes have been applied. Such products are created to meet the needs of the mission team and are not part of the anticipated mission return detailed in the archive or data management plan. Not surprisingly though, these products can be as valuable to future users (in particular to proposers to Data Analysis Programs) as they were to the mission teams.

Some missions offer these products to the PDS for archiving, but this can be done only on a best-effort basis because the products represent additional deliveries beyond what was included in the original plan. These higher-order products generally cannot be archived using the same mission pipeline that produces the required mission archive products because of their unique characteristics and metadata as compared to the raw and calibrated mission data sets.

PDS nodes have found, however, that these higher-order data products are very valuable to end users of these data collections. The node managers, as members of the discipline communities they serve, encourage missions to include these products in the archive and try to assist in their preparation as much as possible. However, this is an unanticipated and unbudgeted effort; and when end-of-mission and end-of-funding approach, these valuable additions to the archive are sometimes lost.

Finding XIII: Higher-order products produced by mission teams beyond what is in their original data management plans are extremely valuable additions to the archive but are not always included due to lack of resources needed by missions to complete the archiving process.

Remediation might include encouragement from NASA management to the mission teams to include higher-order products produced by them in their PDS archive submission so that mission teams can work the extra effort into the delivery schedule, coupled with additional resources to complete the archiving process for these unanticipated but extremely valuable products.

4.7 Laboratory Data and Physical Samples

Finding XIV: Astromaterials Data I

The PDS Level 1 requirements include archiving of mission data, which in principle includes data from laboratory analysis of returned samples from Apollo, Genesis, Stardust, Hayabusa, OSIRIS-

REx, and future missions (Section 3.7). From its collection of >250,000 numbered samples, JSC curation distributes ~1400 samples per year to PIs around the world, with >20,000 samples on loan to >350 PIs in late 2015. Tracking sample descriptions, sample handling and allocation histories, and providing information to PIs represent significant investments in operational data management. However, the NASA Astromaterials Acquisition and Curation Office, in response to RFI NNH 15ZDA012L, reports (Appendix J, note 46) that “JSC curation has **not** made an effort to systematically record the scientific data that has been produced from studies of the samples that it curates.”

JSC curation has initiated a collaboration with the NSF-funded Interdisciplinary Earth Data Alliance (IEDA) to archive data on Apollo samples. They envision terrestrial geochemical databases (e.g., Earthchem, PetDB) as templates to collect astromaterials data that could be cross referenced with and discoverable from the PDS. Within IEDA, MoonDB [28] is funded by NASA through PDART to develop an archive to compile data from Apollo samples, in collaboration with the PDS Cartography and Imaging Sciences Node.

Finding XIV: A large amount of data from laboratory analyses of samples obtained by NASA missions is not archived and is in danger of loss. Astromaterials data today are primarily stored on short-lived media, in private holdings, and with PI-dependent documentation.

A remediation strategy has been initiated by JSC curation; however, it is not clear that the IEDA initiative is PDS-equivalent, as defined in ROSES-2017 (page C7-4). PDS could start a conversation with JSC curation to leverage PDS expertise in ensuring that MoonDB and follow-on efforts are OAIS compliant (thus ensuring interoperability with PDS4). The JSC RFI response (Appendix J) suggests that JSC is apprehensive of the effort required to archive to PDS4 standards, but such a conversation might allay many of these concerns. A new PDS data node in partnership with an existing node (see Section 2.1), or even a full-fledged PDS Node, to archive analytical data collected from extraterrestrial samples returned by NASA missions might be a long-term solution.

Finding XV: Astromaterials Data II

Beyond mission-sourced samples, JSC curation distributes Antarctic meteorites and stratosphere-collected cosmic dust particles, and museums (e.g., NMNH, AMNH, FMNH) distribute non-Antarctic meteorites to PIs. As with mission-sourced samples, chemical, isotopic and other data on extraterrestrial samples are generally collected by individual PIs, across a great many non-standardized instruments and laboratories, presenting significant challenges by comparison with missions, which are mandated to submit specific deliverables to PDS (Section 3.7). Individual analysts are not currently mandated to archive with PDS or any similar system.

Finding XV: A large amount of data from laboratory analyses of meteorites and cosmic dust is not archived and is in danger of loss.

Remediation strategy: The PDS has the opportunity to work in close collaboration with ROSES PIs and others to archive sample-derived data at an individual PI's request, compliant with PDS requirements. ROSES could require PIs to archive data in a PDS4-compliant, long-term repository.

4.8 PDS Structure and Governance

Finding XVI: Node Organization

The PDS node structure (six discipline nodes, two service nodes) has been very stable over the life of the project. Despite NASA being careful to include language in AOs to re-compete the nodes indicating that the existing node structure could be changed in whole or in part, the same nodes have emerged and, in nearly all cases, the existing node teams won the re-compete with little or no real competition. A contributing factor to the small number of competing proposals for existing nodes likely lies in the nearly flat funding profile for PDS generally. Existing nodes have already made significant investment in infrastructure and training. It is very difficult to create a competitive proposal to establish new infrastructure and to train new personnel in addition to maintaining current node services and operations for the same cost as continuing an existing node at its current institution.

Over the course of the last 25 years, though, there has been significant growth in planetary science and in information technology, so it does seem odd that there have been no new PDS nodes—discipline or support nodes—created. This also seems likely due to funding limitation, as existing nodes would be unable to maintain current service levels if new nodes were added with no overall increase in PDS funding.

There are, in fact, new disciplines that could be excellent candidates for new PDS nodes: an Astromaterials and Laboratory Analysis Node, for example, could archive the data produced by lab studies of these targets; or an Exoplanet Node could provide a nexus for collecting observations and analytical results of exoplanet surveys and observations.

Finding XVI: PDS funding levels, combined with the lack of opportunity to propose new nodes separate from the re-compete activity for existing nodes, has had the effect of strongly discouraging the establishment of new nodes or otherwise allowing the PDS organization to grow to keep pace with development and expansion of Planetary Science disciplines and technology.

A suggested remediation would be to include an RFI and CAN specifically targeted at adding new nodes, and thus new funding, to PDS for emerging Planetary Science Disciplines or technical support.

Finding XVII: Transparency

The primary function of the PDS is to preserve and distribute the data that comprise its archive. As part of that function, the PDS has developed standards for formatting and documenting archival science data and associated information. The PDS4 standards build on the cumulative organizational experience of the PDS to define metadata as a detailed information model (IM), which can be expanded and extended for new types of metadata as the various planetary science disciplines grow and develop. The PDS IM addresses both structure and categorization of metadata, as well as metadata needed to provide interoperability between similar, OAIS-based archival organizations. Consequently, a major secondary function of the PDS is to act as a standards development and support organization for its own standards.

Successful open standards organizations, like the World Wide Web Consortium (W3C) or the International Virtual Observatory Alliance (IVOA), encourage and facilitate public engagement in their standards development process through a significant web presence in which they regularly post news and progress updates, distribute documentation and support materials, and invite public comment and participation through collaborative mailing lists and wiki sites (i.e. W3C Standards[42] and IVOA[43]).

In contrast, the PDS standards development process is closed and opaque to its general user community. Only PDS employees can file problem reports or raise development issues; see any sort of tracking or progress reports; or have advance access to test builds, those only for the schema products themselves, not the support tools or documentation. While an open, community-run development effort for PDS standards would not be appropriate (PDS focus needs to remain fixed on NASA's goals for archiving its mission data and maintaining usability over generations.), a greater degree of transparency and additional conduits for public input such as problem reports. This would likely improve PDS relations with the user and developer communities and could, in turn, increase adoption and accommodation of the PDS4 standards.

Finding XVII: The use and application of PDS4 standards and the development of third-party support for PDS4 metadata and formats is hindered by a lack of transparency in the PDS development process.

Remediation would require significant effort. Designing and maintaining a website similar to the sites referenced above to act as a hub for PDS4-related development would require additional personnel and infrastructure support. Participating in and/or monitoring public forums for comment on standards and tools would also require a significant and ongoing investment. Additionally, this online hub could support the dissemination of advanced test builds and documentation, although the preparation of such materials would require support. The costs of these remediations would likely be offset by intangibles such as better community relations and greater acceptance, as well as by more material benefits like third-party tool production and increased access to community expertise and input to design discussions.

[42] https://www.w3c.org/standards/
[43] http://wiki.ivoa.net/twiki/bin/view/IVOA/WebHome

Finding XVIII: PDS Governance

NASA has embarked upon providing more explicit governance and definition to the tasks that the Agency undertakes and how it undertakes them[44]. The PDS is a special case in this evolution, as specifics of its governance have been stated explicitly only in the original research announcement for setting up the node structure (NRA-89-OSSA-1, issued 28 February 1989) and the subsequent proposal calls (NRA 03-OSS-04, issued 1 October 2003 and CAN NNH15ZDA006C[45], issued 5 March 2015), over which time the PDS structure itself has continued to evolve.

The PDS is not currently classified under one of the NASA Procedural Requirement (NPR) rubrics. Three likely possibilities have been identified. NPR 7120.5E describes flight programs and projects[46]; however, although PDS is an adjunct to flight programs within the Planetary Science Division, it would be poorly described as a flight program itself. NPR 7120.7 describes infrastructure programs and projects[47], but also seems a poor fit. The closest descriptor is NPR 7120.8[48], which applies to Research and Technology (R&T) programs and projects. A further difficulty is whether to define PDS as a program or a project. Although PDS is a single-focused activity, like most projects, it does not fit the general definition of a project under NPR 7120.8§2.2 as having "a life-cycle cost, a beginning, and an end," since the mission of PDS is to preserve its data holdings indefinitely into the future. On the other hand, all programs currently defined under NPR 7120.8 contain multiple projects, although there is precedent under NPR 7120.5E for single-project programs. The best solution would be to find a way to define PDS as a single-project program under NPR 7120.8, with the "project", in this case, comprising the selection/funding cycle of the nodes. That is, the PDS "life cycle" is one tied to the funding vehicle, which has a beginning, middle, and end.

PDS is currently organized as a federated data system[49], composed of six competed, PI-led Discipline Nodes funded through NASA Cooperative Agreements, and two Technical Support Nodes under contract at JPL. A Project Office at GSFC and associated NASA Headquarters management directs functionality of the PDS, e.g., in setting meeting agendas, preparing budgets, and distributing funds to the Nodes. Technical Authority[50] is vested in a Management Council (MC) (cf. §1.5 of the Cooperative Agreement Notice at footnote 2), which is composed of the PIs of the Discipline Nodes, the leaders of the Technical Support Nodes, and the Project Manager and Deputy Project Manager of the Project Office. The MC sets policies and determines priorities across PDS as a whole, including those involving interfaces between PDS and missions, other agencies, and NASA itself. The MC collectively sets priorities for development work and schedules for the overall PDS and the individual nodes. The MC operates as a democracy, with a preference for strong consensus.

---

[44] https://nodis3.gsfc.nasa.gov/npg_img/N_PD_1000_000B_/N_PD_1000_000B_.pdf
[45] https://nspires.nasaprs.com/external/solicitations/summary.do?method=init&solId={6C202F60-D0E1-91E6-0D59-A6FC8C96EA36}&path=closedPast
[46] https://nodis3.gsfc.nasa.gov/displayDir.cfm?t=NPR&c=7120&s=5E
[47] https://nodis3.gsfc.nasa.gov/displayDir.cfm?t=NPR&c=7120&s=7
[48] https://nodis3.gsfc.nasa.gov/displayDir.cfm?t=NPR&c=7120&s=8
[49] https://pds.nasa.gov/about/organization.shtml
[50] https://nspires.nasaprs.com/external/viewrepositorydocument/cmdocumentid=449464/solicitationId=%7B6C202F60-D0E1-91E6-0D59-A6FC8C96EA36%7D/viewSolicitationDocument=1/PDSCAN Amend1.pdf

There have, in the past, been changes in the NASA governance approach to PDS as the PDS has evolved to keep pace with technology and user needs. For example, initially there was a Central Node at JPL whose function was, in part, to provide management to run, fund, and coordinate the other PDS Nodes [3]. As information technology evolved, the need for the physical delivery of archived material from the Central Node ended. The complexity and diversity of the mission data sets increased and led to the need for more specialized catalogs, whose maintenance of, and expertise for, lay at the Discipline Nodes. Budgetary pressures and the lack of a demonstrated need of flight-hardware-like, system-engineering approaches for the efficient working of PDS, led to the down-scoping of the Central Node to the support role of the present Engineering Node, the shift from contracts to grants (and now to Cooperative Agreements) for Node support, and the establishment of the independent Project Office, within the Solar System Exploration Data Services Office at GSFC in 2005.

The most recent phase of PDS evolution has tended to increase the demands made on the Project Office. Specifically, the requirements upon the Project Office in budgeting, supporting infrastructure improvements to meet NASA's IT Security requirements, user metrics, and so on, must now be accommodated on a "best effort" basis. Repositioning PDS within the NASA organization, as well as increasing the size of the Project Office to meet these new challenges, may well improve the relationship between NASA and PDS, and put PDS in a better position to present a unified, transparent, and accessible system to its users.

Finding XVIII: NASA management has not settled the question of how PDS fits into current NASA governance structures. PDS has a minimal Project Office, which lacks resources for providing detailed cross-discipline reports, studies, and guidance as there are within other NASA SMD data activities, which would put a more unified public face on the PDS and support other activities provided for in the current NASA governance model.

Remediation: If NASA PSD/SMD management would like to have a more unified response from the PDS nodes in responding to stakeholders (including, e.g., more detailed metrics to document tasks and progress) then an increased Project Office presence needs to be established and funded with new money, i.e., a significantly increased PDS budget. In addition, such increased support and staff would be needed to bring the PDS up to the level of a single-project program under NPR 7120.8, the structure which best describes the governance requirements and functionality of the PDS as currently expected by the Planetary Science Division.

Finding XIX: Timing of the Next PDS Roadmap Study

The goal of this Roadmap Study activity is to "develop a practical, community-developed pathway to implement the new long-term vision for the PDS" (Appendix A). In general, pathways or plans are formulated first, and then implemented. In the case of the PDS, creative solutions are most naturally implemented when the Node re-compete proposals are submitted to NASA, and endorsed when NASA selects a Node. In the case of this particular Study, that competition occurred in 2015, and selections were made. This Roadmap Study was initiated in 2015, and the pathway we develop here may have a lessened impact because Nodes already have 5-year plans, endorsed by NASA, that they have just begun to implement.

Finding XIX: This Roadmap Study was initiated in the year immediately following a recompetition of the PDS Nodes, and will be completed at least three years (and perhaps longer) before the next recompetition, which limits the impact of a Roadmap Study activity on shaping the work of the PDS.

A suggested remediation would be to ensure that the next major Roadmap Study be timed such that its report would be delivered no more than three months, e.g. a fiscal year quarter, before the call for proposals to a PDS Node recompetition.

5 Conclusion / Summary

The Planetary Data System continues to fulfill a vital role in the preservation and curation of data obtained at great expense and with great effort from bodies other than Earth within the solar system. The risk of these data being lost forever, faced some 35 years ago, has been averted. Upgrading PDS to PDS4 over the past decade has further enhanced the possibility of future preservation and use for the long term by securing the data collection to well-known and valued international standards for knowledge preservation and description.

As with any such large endeavor, there have been both known and unknown challenges, exacerbated with sometimes marginal budgets and the overall—and continuing—rapid evolution of IT infrastructure and how it impacts not only the U.S. digital world but the increasingly interconnected and world-wide "digital universe." Such changes, in hardware, software, and middleware[51] continue to evolve. Whether new approaches, such as hyper-convergence[52] or others yet to be defined, will set future IT infrastructure is a question that will be answered by IT innovators dealing with data access, storage, and manipulation on much large scales than of relevance here. Nonetheless, it is in the best interests of the PDS, and indeed of all of NASA's SMD Data Centers, to remain early adopters so as not to be left behind. This requires continued vigilance and investment, both in people and technology.

At the same time, initiatives in massive data analysis [16], colloquially referred to as "big data" is driving a shift of science data system focus from data stewardship toward data-driven knowledge discovery via collaborative analytics and visualization. Innovative technologies are researched and developed by agencies and industries to address challenges presented in this trend. Users increasingly want the ability to find, visualize, and analyze data without having to download and upload. Within NASA's SMD, this approach is already well underway in climate research within

[51] https://en.wikipedia.org/wiki/Middleware
[52] https://www.petri.com/what-is-hyper-convergence

the SMD Earth Science Division via EOSDIS and its components [30][53,54] via the CMR (Section 2.1.1).

Within the PDS, the PDS4 web-service-oriented architecture and its interoperable standards provide an excellent foundation to support more online interactivity to work with, understand, and analyze PDS data by building and integrating analysis and visualization tools built by others. Unlike the CMR in EOSDIS, PDS4 accommodates the less centrally managed PDS node-holdings in order to accommodate a far more diverse target set and even more diverse number of observation types inherent to planetary science observations as compared to Earth science observations.

All of that said, the transition to PDS4 remains incomplete. With many data holdings, and some legacy holdings that are still being collected, in PDS3, the collection is—and may well remain—a hybrid. Transitioning holdings from PDS3 to PDS4 can be labor intensive, and thus dependent upon financial resources. Of more impact is the fact that development of both PDS4 tools and their documentation continues to lag both PDS plans and user-community expectations. There is significant work to do to realize the full potential of the architecture. Meanwhile, the injection of new data sets into PDS4 must remain a priority.

While there has been significant debate among members of the RST about when and/or if such increases in the scope of the PDS are warranted in the future and what they would require in resources, it is clear that the PDS needs to continue to be forward-looking in advancing data usability. Developing and integrating more analysis and visualization tools over time, when resources are available, is one path that will continue to need close scrutiny as information technologies continue to evolve over the coming decade. Such is the position of the early adopter of best IT practices, a critical minimum position for the PDS to remain within for the decade to come.

[53] http://eijournal.com/print/articles/managing-big-data
[54] https://earthdata.nasa.gov/getting-petabytes-to-people-how-the-eosdis-facilitates-earth-observing-data-discovery-and-use

6 Acknowledgement

The Roadmap Study Team appreciates the many responses to the Request for Information (NNH-15ZDA012L) that NASA solicited. Their comments were included in our deliberations.

7 Acronyms

Acronym	**Definition**
ACM	Association of Computing Machines
ADASS	Astrophysical Data Analysis and Software Systems
ADS	Astrophysics Data System
AGU	American Geophysical Union
AIAA	American Institute Aeronautics and Astronautics
AMNH	American Museums of Natural History
ANNEX	Appendix
AO	Announcement of Opportunity
API	Application programming interface
APL	Applied Physics Lab (John Hopkins University)
APUS	American Public University System
ARC	Ames Research Center
ASEE	American Society for Engineering Education
ASME	American Society of Mechanical Engineers
ASU	Arizona State University
ATM	Atmosphere Node (PDS)
CAN	Cooperative Agreement Notice
CAPTEM	Curation and Analysis Planning Team for Extraterrestrial Materials
CDF	Common data format
CMR	Common Metadata Repository
CNSA	China National Space Agency
CODMAC	Committee on Data Management and Computation
COMET	Configuration Management EOSDIS Tool
COSPAR	International Committee on Space Research
CSV	Comma Separated Variable
DAAC	Distributed Active Archive Center
DAP	Data Analysis Program
DLR	German National Aerospace Center
DMP	Data Management Plan
DNA	DeoxyriboNucleic Acid
DOI	Digital Object Identifier
DPS	Division of Planetary Science
DSN	Deep Space Network
DVD	Digital Video Disk
EDOS	EOS Data and Operations System
EMS	ESDIS Metric System
EOS	Earth Observing System
EOSDIS	Earth Observing System Data and Information System
ESA	European Space Agency
ESDIS	Earth Science Data and Information System
ESMO	Earth Science Mission Operations
EUMETSAT	EUropean Organization for the exploitation of METeorogical Satellite

FACA	Federal Advisory Committee Act
FITS	Flexible Image Transport System
FMNH	Field Museum of Natural History
FWG	FITS Working Group
GCMD	Global Change Master Directory
GDAL	Geospatial Data Abstracting Library
GEO	Geosciences Node
GIBS	Global Imagery Browse Services
GIF	Graphics Interchange Format
GSFC	NASA Goddard Space Flight Center
HDF	Hierarchal Data Format
HDN	HIRISE Data Node
HEASARC	High Energy Astrophysics Science Archive Research Center
HPD	Heliophysics Division (NASA)
HST	Hubble Space Telescope
IASF	Italian Astrophysics Space Facility
IAU	International Astronomical Union
ICD	Interface Control Document
IDL	Interactive Data Language
IEDA	Interdisciplinary Earth Data Alliance
IEEE	Institute Electrical and Electronic Engineering
IMG	Cartography and Imaging Science (CIS) Node
INAF	Italian National Institute for Astrophysics
IPAC	Infrared Processing and Analysis Center
IPDA	International Planetary Data Alliance
IRSA	InfraRed Science Archive
ISBN	International Standard Book Number
ISIS	Integrated Software for Images and Spectrometers
ISO	International Standards Organization
ISRO	Indian Space Research Organization
ISS	International Space Station
ITAR	International Traffic in Arms Regulations
JAXA	Japanese Aerospace eXploration Agency
JHU	Johns Hopkins University
JHUAPL	Johns Hopkins University Applied Physics Laboratory
JIRAM	Jovian Infrared Auroral Mapper
JPEG	Joint Photographic Experts Group
JPL	NASA’s Jet Propulsion Laboratory
JSC	NASA Johnson Space Center
JWST	James Webb Space Telescope
LADEE	Lunar Atmosphere and Dust Environment Explorer
LAMBDA	Legacy Archive for Microwave Background Data Analysis
LANCE	Land, Atmosphere Near real-time Capability for Earth
LAPLACE	NASA/ESA proposed mission to Europa
LDEX	Lunar Dust EXperiment
LDN	Lunar Data Node (PDS)
LDP	Lunar Data Project (GSFC)
LGP	Lunar Geology and Petrology (data node)
LGPDN	Lunar Geology and Petrology Data Node
LLC	Limited Liability Company
LONEOS	Lowell Observatory Near Earth Object Search
LPSC	Lunar Planetary Science Conference
LROC	Lunar Reconnaissance Orbiter Camera

MAP	Map a Planet
MAP2	Map a Planet 2
MAPSIT	Mapping and Planetary Spatial Infrastructure Team
MAST	Mikluski Archive Space Telescope
MAVEN	Mars Atmosphere and Volatile EvolutioN
MESSENGER	MErcury Surface, Space ENvironment, GEochemistry, and Ranging
MIC	Micro-particle Impact Collision
MOU	Memorandum of Understanding
MSFC	Marshall Space Flight Center
NAIF	Navigation and Ancillary Information Facility
NAS	National Academy of Sciences
NASA	National Aeronautics and Space Administration
NEAR	Near Earth Asteroid Rendezvous
NEAT	Near Earth Asteroid Telescope
NEO	Near Earth Object
NMNH	National Museum of Natural History
NMSU	New Mexico State University
NRA	NASA Research Announcement
NRC	National Research Council
NSF	National Science Foundation
NSSDC	National Space Science Data Center
NSSDCA	NASA Space Science Data Coordinated Archive
OAIS	Open Archival Information System
ODE	Orbital Dynamic Explorer
OLAF	On-Line Archive Facility
OPUS	Outer Planet Universal (data) Search
OSIRIS-REx	Origins, Spectral Interpretation, Resource Identification, Security, Regolith Explorer
OSSA	Office Space Science and Applications
OSTI	Office of Scientific and Technical Information
OSTP	Office of Science and Technology Policy
PAC	Planetary Advisory Committee (NASA)
PDAP	Planetary Data Analysis Program
PDART	Planetary Data Archiving, Restoration, and Tools
PDF	Portable Document Format
PDMAP	Planetary Data Management and Archive Plan
PDS	Planetary Data System
PDSMC	PDS Management Council
PHOENIX	NASA Mars lander mission
PILOT	Planetary Image LOcator Tool
PLOS	Public Library of Science
POC	Point of Contact
POW	Curve plotting tool from HEASARC
PPI	Planetary Plasma Interactions Node (PDS)
PSA	Planetary Science Archive (ESA)
PSD	Planetary Science Division (NASA)
PSI	Planetary Science Institute
QQC	Quality, quantity, and continuity
REST	Representational State Transfer
RFI	Request for Information
RFP	Request for Proposal
ROSES	Research Opportunities in Space and Earth Sciences
RST	Roadmap Study Team
SAE	Society of Automotive Engineers

SBAG	Small Bodies Assessment Group
SBIB	Small Bodies Image Browser
SBMT	Small Bodies Mapping Tool
SBN	Small Bodies Node (PDS)
SDAC	Solar Data Analysis Center
SDPS	Science Data Processing System
SELENE	SELenological and ENgineering Explorer "KAGUYA"
SETI	Search for Extra Terrestrial Intelligence
SIPS	Science Investigator-led Processing Systems (EOS)
SIS	Software Interface Specification
SMD	Science Mission Directorate (NASA)
SPDF	Space Physics Data Facility
SPICE	Spacecraft, Planet, Instrument, C-matrix, and Events
SSB	Space Science Board
SSED	Solar System Exploration Division
STEM	Space, Technology, Engineering, and Mathematics
TDN	Temporal Dependence Network
THEMIS	Time History of Events and Macroscale Interactions during Substorms
UCLA	University of California at Los Angeles
UMD	University of MarylanD
UMM	Unified Metadata Model
UPC	Unified Planetary Coordinates
URL	Uniform Resource Locator
USGS	United States Geological Survey
VESPA	Virtual European Solar and Planetary Access
VICAR	Video Image Communication and Retrieval
VIMS	Visible and Infrared Mapping Spectrometer
VIR	Visible Infrared Spectrometer
VIRTIS	Visible InfraRed Thermal Imaging Spectrometer
XML	eXtensible Markup Language

8 References

[1] R. Bernstein, S.S. Alexander, R. Arvidson, O. Ely, C. Goad, E.J. Groth, D.A. Landgrebe, R. Legeckis, R.L. McPherron, H. Tananbaum, D.L. Turcotte, T.H. Vonder Haar, A.G.W. Cameron, M.A. Chinnery, R.C. Hart, Data Management and Computation: Volume 1: Issues and Recommendations, National Academy Press, Washington, D.C., 1982.

[2] S.W. Lee, The Planetary Data System, Reviews of Geophysics Supplement, U.S. National Report to International Union of Geodesy and Geophysics, 1987-1990, 29 (1991) 337-340.

[3] S.K. McMahon, Overview of the Planetary Data System, Planetary and Space Science, 44 (1996) 3-12.

[4] T.H. Morgan, R.L. McNutt, Jr., Renewing the Planetary Data System - Roadmapping the needs of the community 2017 - 2026, in: 47th Lunar and Planetary Science Conference, The Woodlands, Texas, 2016, pp. 2.

[5] R.L. McNutt Jr, A historical study of payload masses to Earth orbit and beyond, in: 46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, AIAA, Nashville, TN, 2010, pp. 7.

[6] J. McDowell, Launchlog master list, in, 2017.

[7] R.E. Arvidson, G. Hunt, D. Landgrebe, E. Levinthal, G.H. Ludwig, T.B. McCord, E. Schreier, R. Walker, G. Wiederhold, Issues and recommendations associated with distributed computation and data management systems for the space sciences, National Academy Press, Washington, D.C., 1986.

[8] C.T. Russell, M. Abbott, T. Albert, L.K. Bolef, S. Brecht, J.E. Estes, G. Fabbiano, O.K. Garriott, J.C. Gille, E. Hansen, H. Masursky, L. McFadden, N. Roussopoulos, P. Shames, D.N. Baker, R.C. Hart, C.J. Chamberlain, Selected issues in space science data management and computation, National Academy Press, Washington, D.C., 1988.

[9] R.E. Arvidson, The NASA Planetary Data System: Past, present, and future, in: Second Planetary Data Workshop, Flagstaff, Arizona, 2015, pp. 2.

[10] E.A. Guinness, R.E. Arvidson, S. Slavney, The Planetary Data System Geosciences Node, Planetary and Space Science, 44 (1996) 13-22.

[11] E.J. Grayzeck Jr, M.F. A'Hearn, A.C. Raugh, M.V. Sykes, D.R. Davis, D. Tholen, Services of the Small Bodies Node of the Planetary Data System, Planetary and Space Science, 44 (1996) 47-54.

[12] R.J. Walker, S.P. Joy, T.A. King, C.T. Russell, R.L. McPherron, W.S. Kurth, The Planetary Plasma Interactions Node of the Planetary Data System, Planetary and Space Science, 44 (1996) 55-64.

[13] M.R. Showalter, K.J. Bollinger, J.N. Cuzzi, P.D. Nicholson, The Rings Node for the Planetary Data System, Planetary and Space Science, 44 (1996) 33-45.

[14] E.M. Eliason, S.K. LaVoie, L.A. Soderblom, The Imaging Node for the Planetary Data System, Planetary and Space Science, 44 (1996) 23-32.

[15] C.H. Acton Jr, Ancillary data services of NASA's Navigation and Ancillary Information Facility, Planetary and Space Science, 44 (1996) 65-70.

[16] M.I. Jordan, K.M. Carley, R.D. Coifman, D.J. Crichton, M.J. Franklin, A.C. Gilbert, A.G. Gray, T.J. Hastie, P. Indyk, T. Johnson, D. Lambert, D. Madigan, M.W. Mahoney, E.M. Maley, C. Olston, Y. Singer, A.S. Szalay, T. Zhang, Frontiers in massive data analysis, in, The National Academies of Sciences, Engineering, and Medicine, Washington, D.C., 2013, pp. 190.

[17] Z.D. Stephens, S.Y. Lee, F. Faghri, R.H. Campbell, C. Zhai, M.J. Efron, R. Iyer, M.C. Schatz, S. Sinha, G.E. Robinson, Big Data: Astronomical or Genomical?, PLOS Biology, 13 (2015) e1002195.

[18] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E.M. LeProust, B. Sipos, E. Birney, Towards practical, high-capacity, low-maintenance information storage in synthesized DNA, Nature, 494 (2013) 77-80.

[19] A. Extance, How DNA could store all the world’s data, Nature, 537 (2016) 22-24.

[20] A.C. Raugh, J.S. Hughes, Metadata wrangling in the new PDS4 standards, in: 2015 Astronomical Data Analysis Systems and Software Conference ADASS XXV, Astronomical Society of the Pacific Sydney, New South Wales, Australia, 2015, pp. 4.

[21] L. Gaddis, T. Hare, Status of tools and data for planetary research, in: EOS Earth and Space News, American Geophysical Union, 2015, pp. 1.

[22] A.C. Raugh, J.S. Hughes, An overview of PDS4 archival file formats, in: A.R. Taylor, E. Rosolowsky (Eds.) Astronomical Data Analysis Software and Systems XXIV (ADASS XXIV), Astronomical Society of the Pacific, Calgary, Alberta, Canada, 2014.

[23] J.S. Hughes, R.R. Downs, An assessment of science discipline archive against ISO 16363, in: 2016 American Geophysical Union Fall Meeting, San Francisco, California, 2016, pp. 1.

[24] J.S. Hughes, D. Crichton, R. Simpson, M. Gordon, R. Joyner, A. Raugh, E. Guiness, M. Martin, The PDS4 information model design principles - How well did they work?, in: PV 2015 Conference, EUMETSAT, Darmstadt, Germany, 2015, pp. 8.

[25] J.S. Hughes, D. Crichton, S. Hardman, E. Law, R. Joyner, P. Ramirez, PDS4: A model-driven planetary science data architecture for long-term preservation, in: 2014 IEEE 30th International Conference on Data Engineering Workshops, 2014, pp. 134-141.

[26] D. Crichton, J.S. Hughes, S. Hardman, E. Law, R. Beebe, T. Morgan, E. Grayzeck, A Scalable Planetary Science Information Architecture for Big Science Data, in: 2014 IEEE 10th International Conference on e-Science, 2014, pp. 196-203.

[27] J.S. Hughes, D.J. Crichton, C.A. Mattmann, Ontology-based information model development for science information reuse and integration, in: 2009 IEEE International Conference on Information Reuse & Integration, 2009, pp. 79-84.

[28] K.A. Lehnert, C. Evans, N. Todd, R. Zeigler, Restoration and synthesis of lunar geochemical and petrological sample data to support future science (MoonDB), in: Second Planetary Data Workshop, Flagstaff, Arizona, 2015, pp. 2.

[29] J. Rahe, The NASA Planetary Data System, Highlights of Astronomy, 9 (1992) 717-718.

[30] NASA, EOSDIS Handbook, Version 1.3 ed., NASA, 2017.

Terms of Reference for the Planetary Data System Roadmap Team

BACKGROUND

NASA's planetary missions archive their data in the Planetary Data System (PDS). The PDS has recently (2015) completed a full and open competition for the Discipline Nodes within the PDS. The PDS has also just completed (2016) a performance review of the two technical support elements within the PDS architecture: the Engineering node (EN) and the Navigation and Ancillary Data Facility (NAIF).

PURPOSE

The NASA PDS Roadmap team will complete a PDS Roadmap for the period 2017-2026. The purpose of this activity is to provide a forecast of both the rapidly changing information technology (IT) environment and the changing expectations of science communities with respect to Planetary Data archives. The Roadmap Team will report their findings to the PSD Director.

APPONTMENT, STAFFING, SCHEDULE, AND REPORTING

The majority of the PDS Roadmap Team will consist of 10 to 12 members chosen from those who have proposed to the "Dear Colleague Letter to Solicit Nominations for the NASA Planetary Data System (PDS) Roadmap Team". However, NASA reserves the right to appoint additional members as needed. In particular, the PDS Chief Scientist, Dr. Ralph L. McNutt, Jr. (Johns Hopkins University Applied Physics Laboratory, JHUAPL) will chair the meetings of this group, and Ms. Emily Law of the PDS Engineering Node will serve as his alternate as needed. The PDS Project Manager Dr. Thomas H. Morgan (Goddard Space Flight Center, GSFC) is an *ex officio* team member.

The PDS Chief Scientist will be responsible for the timing and agenda (with the concurrence of the PDS Project Manager) of each meeting. The team may hold town-hall meetings at national science meetings or survey the provider community by email and Web poll. The Chief Scientist will report the results of the Roadmap Activity to the Planetary Science Division Director Dr. James L. Green at the end of one year of effort.

Team members will be identified and contacted in late March 2016. The Chief Scientist will set up the first virtual meeting in mid-April 2016. At the first meeting the Chief Scientist will solicit input on a Roadmap outline, engage the team in discussion of the relative importance of future partnerships and directions of PDS, and work to understand member schedules to plan future meetings. The team will provide a mid-term report to NASA HQ in August 2016 and a final report by the end of November 2016.
The Roadmap Team is independent of the PDS Management Council. 11 March 2016

[55] https://pds.nasa.gov/roadmap/Term%20of%20Reference%20.pdf

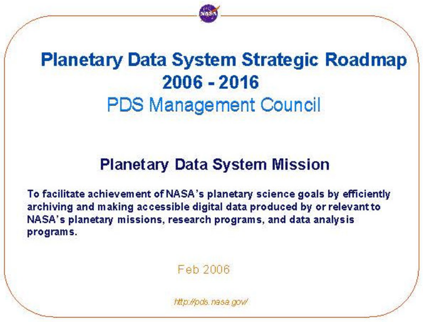
Roadmap Study Team Members

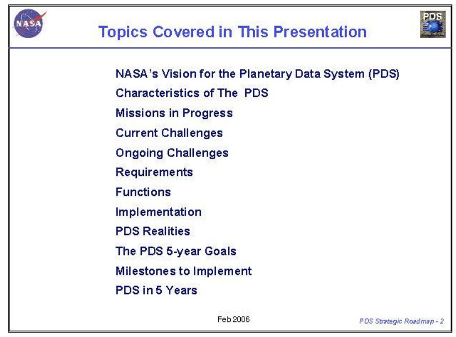
To move forward with this effort, NASA called for self-nominations from the planetary science community for a Roadmap Study Team (RST) (https://pds.nasa.gov/pdsroadmapteam.shtml). RST members included:

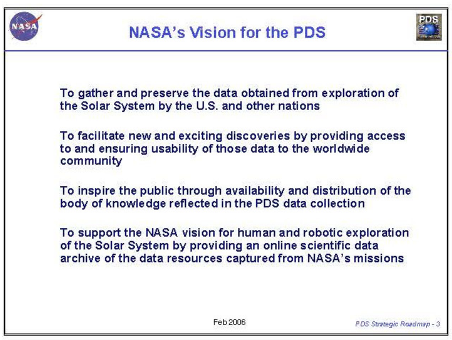
Name	**Affiliation**
Amitahba Ghosh	Tharsis, Inc.
Anne Raugh	University of Maryland, College Park
Denton Ebel	American Museum of Natural History
Edwin Grayzeck	Cornell Technical Services
Emily Law (Vice Chair)	Jet Propulsion Laboratory
Flora Paganelli	SETI Institute
Katherine Crombie	Indigo Information Services, LLC
Lisa Gaddis (Vice Chair)	United States Geological Survey/Flagstaff
Matthew Tiscareno	SETI Institute
Ralph McNutt (Chair)	Johns Hopkins University Applied Physics Laboratory
Renee Weber	NASA Marshall Space Flight Center
Ross Beyer	SETI Institute and NASA's Ames Research Center
Thomas Stein	Washington University
Thomas Morgan	NASA's Goddard Space Flight Center
Kathryn Powell (Secretary)	Washington University

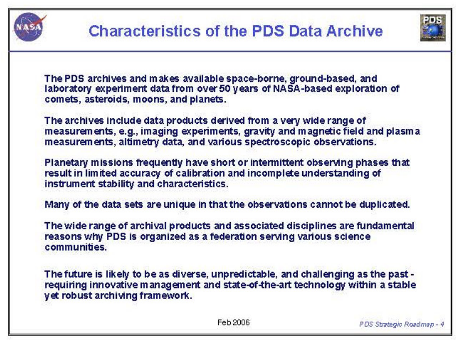
[56] https://pds.nasa.gov/roadmap/index.shtml

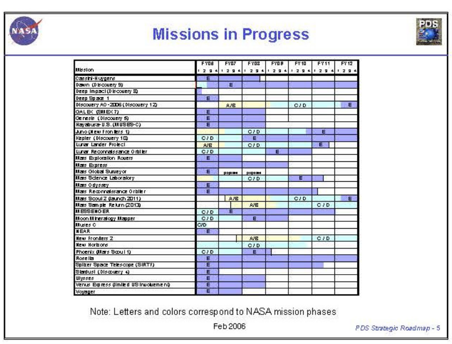
Appendix C. 2006 PDS Strategic Roadmap[57]





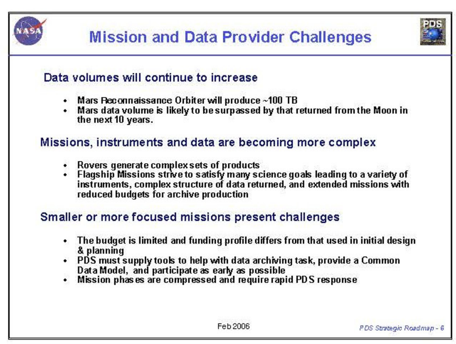






Mission	FY06	FY07	FY08	FY09	FY10	FY11	FY12
Cassini-Huygens	E						
Dawn (Discovery 9)		E					
Deep Impact (Discovery 8)							
Deep Space 1	E						
Discovery AO - 2006 (Discovery 12)		A/B			C / D		E
GALEX (SMEX 7)	E						
Genesis (Discovery 5)	E						
Hayabusa-U.S. (MUSES-C)	E						
Juno (New Frontiers 1)			C / D			E	
Kepler (Discovery 10)	C / D		E				
Lunar Lander Project	A/B		C / D			E	
Lunar Reconnaissance Orbiter	C / D			E			
Mars Exploration Rovers	E						
Mars Express							
Mars Global Surveyor	E	propose	propose				
Mars Science Laboratory			C / D		E		
Mars Odyssey	E						
Mars Reconnaissance Orbiter	E						
Mars Scout 2 (launch 2011)		A/B			C / D		E
Mars Sample Return (2013)			A/B			C / D	
MESSENGER	C / D	E					
Moon Mineralogy Mapper	C / D		E				
Muses C	C/D						
NEAR	E						
New Frontiers 2			A/B			C / D	
New Horizons			C / D				
Phoenix (Mars Scout 1)	C / D		E				
Rosetta	E						
Spitzer Space Telescope (SIRTF)	E						
Stardust (Discovery 4)	E						
Ulysses	E						
Venus Express (limited US involvement)	E						
Voyager	E						

Note: Letters and colors correspond to NASA mission phases



57 https://pds.nasa.gov/roadmap/PDS%20ROADMAP%20Feb%202-2.pdf

User and Customer Challenges

User desires and expectations are becoming more complex and demanding with time as data and mission complexity increase

- Users want and need derived products
- Users expect rapid access, flexible search capability and refined catalogues
- Users want readily accessible data visualization, analysis, and cross-correlation tools
- Experienced users frequently want interfaces with familiar data manipulation packages
- Inexperienced users are "Google" oriented but PDS data is not always adaptable to the expected indexing .

Interdisciplinary Science requires parametric, geographic or spatial searches across missions, instruments and data sets

Lack of reduced data records, combined with less than adequate search and retrieval capabilities associated with the current implementation of PDS sometimes frustrate users who wish to find and download highly processed data for further analyses

Feb 2006 — *PDS Strategic Roadmap - 7*

Challenges Associated with International Collaboration

Planetary exploration has become increasingly international in scope.

- International missions providing the data
- International scientists using the data

The science community expects seamless access to integrated planetary data archives across international boundaries which requires

- Plans and agreements for sharing data
- International data standards
- International protocols for sharing data

Currently, there are no international data standards for planetary science

- PDS has made progress in having its standards become internationally accepted, e.g., ESA's Huygens, Mars and Venus Express data sets will be compliant with PDS.
- The Planetary Science Archive (PSA) and PDS are working on establishing such a standard based on PDS

Coordination and development and application of standards for archive production across international partners substantially increases workload due to increasing number of interfaces and standards training

ITAR issues that may preclude fully open discussion among international collaborators

Feb 2006 — *PDS Strategic Roadmap - 8*

Operational Challenges

Users require that PDS supports heterogeneity for both technology and science

- They need software tools that can run on a wide variety of platforms and adapt for different uses (E.g., different mission and instrument teams)
- They need to be able to use PDS for discipline-specific science both in archiving and searching and retrieving data

Users expect to be able to search all of the PDS data holdings within the system and perform correlated searches for products across nodes

Users expect online access and distribution of all the data holdings

- Users need the ability to download data products and volumes of increasing sizes
- PDS needs the ability to move large volumes between institutions

Upcoming missions require that PDS be able to scale critical functions

- The storage infrastructure must scale to support the increase of volume
- The architecture must scale to support the complexity of missions flying many different types of instruments
- The operations of PDS must scale to support the increasing number of missions that are occurring concurrently

Aging and obsolete technology must be refreshed to ensure long term usability of the data and the system

Data in the archive must be periodically validated to ensure long term preservation

Feb 2006 — *PDS Strategic Roadmap - 9*

Ongoing Challenges

Within a "flat" funding profile, PDS must:

- Interface with new missions (typically more than 20 at one time in various development phases)
- Ingest data from increasingly complex instruments
- Add non-mission data from individuals, labs and observatories
- Incorporate new storage/distribution technologies
- Educate data providers
- Provide useful links to other, related data systems
- Respond to an increasing spectrum of demands from an ever-growing community of users

Feb 2006 — *PDS Strategic Roadmap - 10*

Requirements

PDS will provide expertise to guide and assist missions, programs, and individuals to organize and document digital data that can be used to support NASA's goals in planetary science and Solar System exploration

PDS will collect suitably organized and well-documented data into archives that are peer reviewed and maintained by members of the scientific community.

PDS will make these data accessible to users seeking to achieve NASA's goals for exploration and science

PDS will ensure the long-term preservation of the data and maintain their usability

See backup slides 1-3 for details

Feb 2006 — *PDS Strategic Roadmap - 11*

PDS Functions

PDS serves the scientific community by assuring the availability of high quality and scientifically useful data products

To accomplish these goals, PDS

- works with Data Providers to Prepare Archival Quality Data Products
- provides Access to Data from NASA and International Missions
- delivers Data to the Scientific Community
- establishes a Common Data Model and Data Dictionary for Planetary Data
- sets Archival Standards
- preserves the Data
- assists Scientists in Accessing and Using Planetary Data
- is responsive to a diverse community of users
- facilitates Education and Public Outreach

Feb 2006 — *PDS Strategic Roadmap - 12*

Implementation - Operation

PDS is a close federation Nodes with both Science and Support functions

- Science functions are organized by discipline and include
  - data ingestion
  - data distribution
  - interfacing with data suppliers and users to ensure that
    - maximum science value is captured within the archive
    - the archive is of greatest utility to both immediate and long-term science users
      - Immediate users, by their use of the system, help PDS understand if the services and data sets are of optimal use to the community. These users have the benefit of an active instrument team to whom comments and replies can be passed, if needed, allowing the archive to be modified.
      - Long-term users need final, stand-alone archives because the instrument experts may no longer be available.

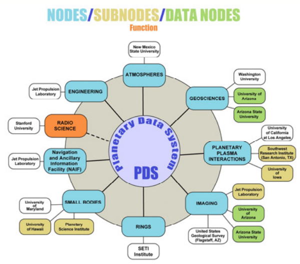
Feb 2006 PDS Strategic Roadmap - 13

Implementation - Operation (cont.)

- Support functions include basic development and cross-discipline support such as
  - common tools, libraries, procedures, and standards for data preparation, submission, and management
  - common tools for data manipulation
  - an infrastructure that facilitates
    - easy navigation within and access to holdings throughout the federation
    - simple system-wide maintenance and upgrades
- The Engineering Node focuses primarily on Support; the other Nodes focus primarily on Science, but all have at least test bed and review tasks

Feb 2006 PDS Strategic Roadmap - 14

Implementation - The PDS System



Feb 2006 PDS Strategic Roadmap - 15

Implementation - Education and Public Outreach

All tools and holdings of the PDS are available to everyone, but the primary customers are scientists and their upper division and graduate students and associates.

PDS Nodes maintain tools that support advanced class work and research projects.

PDS directs users to other valuable related resources

PDS evaluates and accepts requests for product generation and delivery for EPO Projects

Backup sides 4-5

Feb 2006 PDS Strategic Roadmap - 16

Implementation - Assessment and Evaluation

External community assessment is often based on user expectations that do not match requirements and charter

- Well defined criteria for evaluation must be implemented
- Defensible metrics need to be presented

The PDS will maintain an ongoing internal evaluation addressing effectiveness of all PDS functions. Data to support the evaluation process will be assimilated from:

- Regular consultation with individual Node Advisory Groups
- Periodic surveys of the data provider and data user communities
- Unsolicited user feedback
- Storage, access, and retrieval statistics

Feb 2006 PDS Strategic Roadmap - 17

PDS Realities

Inadequate Mission budgets for archiving have caused problems for PDS since its inception

- Data archiving in general has been relatively low-priority for many missions. Increased costs and unexpected mission problems have often pushed archiving further down the priority list.
- Instrument teams have been limited by time and funds in their ability to generate, validate, and deliver to the PDS derived products of use to the science community
- Most missions did not plan adequate funding for development and delivery of derived products.
- Data analysis programs have not been funded to generate, validate, and deliver-compliant derived products to the archive

Feb 2006 PDS Strategic Roadmap - 18

The PDS 5-Year Goals

Significantly improved user services based in part on periodic user feedback, data searches correlated across multiple data sets, new data visualization tools and enhanced delivery capabilities

Each NASA planetary mission plan end-to-end archive generation and achieve on-time delivery of documented and validated PDS-compliant archives

Generation, validation, and delivery of PDS-compliant derived products by NASA mission instrument teams and PI's from data analysis programs as a standard practice

Establish and maintain coordination with domestic and international counterparts to assure interoperability making available data from all planetary missions (i.e., NASA, ESA, and other agencies)

Feb 2006

PDS Strategic Roadmap - 19

Milestones to implement 5-Year Goals

Data Access and Distribution

- PDS will support a reduced set of file structures/formats for science data, with the selection based on community input.
- All Discipline Node catalogs will provide search capabilities across all their holdings at a fine level of granularity, including geometric constraints.
- Basic data visualization tools will be integrated into the search process, so that most data products can be evaluated quickly by the user before selection for delivery.
- Automatic translation tools will allow users to receive data in whatever suitable format they prefer.
- Users will have the option of receiving data requests in the form of fully PDS-compliant custom volumes generated on-the-fly.

Feb 2006

PDS Strategic Roadmap - 20

Milestones to implement 5-Year Goals (cont.)

Team Interfaces & Archiving

- Mission archiving with PDS will be routine
  - Well established milestones that evolve into place based on PDS and community experience and inputs
  - Discipline Nodes will provide tools to the teams for generating the necessary metadata relevant to their discipline
  - A complete set of automated tools will be available to validate the metadata against PDS standards
- AOs need to require all instrument/mission teams to
  - deliver fully calibrated, geometrically corrected data throughout the mission and to resubmit the data when they are improved
  - collaborate with the PDS on all data calibration/processing/retrieval systems designed for their team members. Data systems used by team members must seamlessly transition over to the PDS as the team's data sets become publicly available.

Feb 2006

PDS Strategic Roadmap - 21

Milestones to implement 5-Year Goals (cont.)

Technical Infrastructure

- PDS architecture will be entirely distributed. Facilities at each Discipline Node will be capable of handling cross-discipline queries
- All data holdings will reside at a minimum of two different, geographically separated locations. Popular data sets will be continuously on line at two or more locations
- All PDS software will be written in a manner that minimizes, and preferably eliminates, hardware and OS dependencies
- PDS will upgrade its network connections regularly to provide fast delivery of data to users
- PDS will collaborate with other NASA entities to better assess and utilize existing and emerging storage technologies

To achieve these goals PDS will continue and enhance its proactive role in working with NASA Headquarters and Mission management and instrument teams to identify and maintain necessary processes, schedules, data models, standards and assistance.

Feb 2006

PDS Strategic Roadmap - 22

The PDS in 5 Years

PDS is committed to a major, structured modernization of its data archiving & distribution

The framework will be PDS-4; key aspects of which will include:

- Inputs from the mission & science user communities
- A simplified set of data formats
- Metadata requirements to support modern search technology
- Highly automated validation & ingestion
- On-the-fly conversions from stable, streamlined archival formats to suitable, user requested formats
- Improved interoperability with other domestic and international space data systems

The pace and fidelity of this development will depend on funding levels and headquarters support through the inclusion of specific requirements in future AOs

Feb 2006

PDS Strategic Roadmap - 23

Backup Materials

Feb 2006

PDS Strategic Roadmap - 24

PDS Level 1 & Level 2 Requirements

PDS will provide expertise to guide and assist missions, programs, and individuals to organize and document digital data supporting NASA's goals in planetary science and Solar System exploration

- **1.1** **Single Point of Contact:** PDS will provide a single point of contact to each mission, program, agency or individual (i.e., data providers) wishing to submit archival data
- **1.2** **Expert Help:** PDS will provide expert help in designing archival data sets
- **1.3** **Plans and Documents:** PDS will assist data providers in developing archive plans, interface control documents, validation procedures, and delivery schedules
- **1.4** **Archiving Standards:** PDS will develop, publish, and maintain archiving standards for planetary science data
- **1.5** **Archiving Tools:** PDS will develop, distribute, and maintain tools to assist data producers in assembling, validating, and submitting archival products

Will be updated as new versions are developed

Feb 2006 — *PDS Strategic Roadmap - 25*

PDS Level 1 & Level 2 Requirements (cont.)

PDS will collect suitably organized and well-documented data into archives that are peer reviewed and maintained by members of the scientific community.

- **2.1** **Receive:** PDS will develop, publish, and implement procedures for receiving, acknowledging and tracking data submissions
- **2.2** **Validation:** PDS will develop, publish, and implement procedures for validating data submissions to ensure compliance with standards
- **2.3** **Peer Review:** PDS will develop, publish, and implement procedures for conducting peer reviews of all data submissions to help ensure completeness, accuracy and usability of content
- **2.4** **Acceptance:** PDS will develop, publish, and implement procedures for accepting or rejecting archival data
- **2.5** **Catalog:** PDS will maintain a catalog of accepted archival data sets
- **2.6** **Storage:** PDS will develop and maintain appropriate storage for its archive

Feb 2006 — *PDS Strategic Roadmap - 26*

PDS Level 1 & Level 2 Requirements (cont.)

PDS will make these data accessible to users seeking to achieve NASA's goals for exploration and science

- **3.1** **Search:** PDS will allow and support searches of its archival holdings
- **3.2** **Retrieval:** PDS will facilitate transfers of its data to users

PDS will ensure the long-term preservation of the data and maintain their usability

- **4.1** **Long-Term Preservation:** PDS will determine requirements for and ensure long-term preservation of the data
- **4.2** **Long-Term Usability:** PDS will establish long-term usability requirements and implement procedures for meeting them

Feb 2006 — *PDS Strategic Roadmap - 27*

Education and Public Outreach - Tools

Current PDS tools support advanced class work and research projects.

- MER Analyst's Notebook (http://anserver1.eprsl.wustl.edu/) with data collection, viewable images and analysis notes by and for scientists
- Encyclopedia of Planetary terms (http://atmos.nmsu.edu/)
- Planet viewer, moon trackers and ephemeris generators (http://pds-rings.seti.org/tools/) to use for planning observations, writing proposals and preparing for public nights at educational facilities.
- NASAView (http://pds.jpl.nasa.gov/tools/software_download.cfm) for viewing imaging data
- Planetary Image Atlas (http://pds-imaging.jpl.nasa.gov/Atlas/) for assessing imaging data
- Map-a-Planet (http://pdsmaps.wr.usgs.gov/maps.html/) for selecting specific area maps
- PDS documents page (http://pds.jpl.nasa.gov/documents/) for tutorials on PDS functions
- Comet modeling tools http://pds-smallbodies.astro.umd.edu/comettools/index.html

Feb 2006 — *PDS Strategic Roadmap - 28*

Education and Public Outreach- Related Sites

PDS directs users to other valuable NASA resources

- The Planetary Photojournal (http://photojournal.jpl.nasa.gov/) with accessible images and accurate captions provides a timely resource for students and educators
- The NASA Astrophysical Data System (http://adswww.harvard.edu/) that provides a detailed abstract search
- Imaging and Geosciences nodes provide primary mission interfaces for NASA Regional Planetary Information Facilities (RPIFs) (http://www.lpi.usra.edu/library/RPIF/)
- ISIS web site (http://isis.astrogeology.usgs.gov/) widely used USGS developed software for processing image arrays and cubes
- NASA/USGS Planetary GIS Web Server (http://webgis.wr.usgs.gov/)
- Gazetteer of Planetary Nomenclature (http://planetarynames.wr.usgs.gov/) generated by the International Astronomical Union
- Orbital data for small bodies at the IAU's Minor Planet Center at CFA (largely NASA funded). http://cfa-www.harvard.edu/cfa/ps/mpc
- JPL's Horizons system for orbital data and ephemeredes and to JPL's orbit visualization tool. http://ssd.jpl.nasa.gov/
- A variety of sites for Near-Earth Asteroids, including sites at JPL. http://neo.jpl.nasa.gov/ . http://neat.jpl.nasa.gov/
- A tool for finding comet data in the HST data base http://pdssbn.astro.umd.edu/sbnhtml/comets/HST/

Feb 2006 — *PDS Strategic Roadmap - 29*

In 2016, the PDS Nodes were asked to submit self-assessing comments in reference to the 2006 PDS Strategic Roadmap (Appendix C). In this appendix, **bold** denotes text from the 2006 Roadmap, and plain text denotes the RST's summary of the Nodes' self-assessments.

Five-Year Goals

1 - Significantly improved user services based in part on periodic user feedback, data searches correlated across multiple data sets, new data visualization tools and enhanced delivery capabilities.

The PDS Nodes felt that progress had been made on this goal but acknowledged that they were not where they would like to be. Many of the Nodes pointed to tools that have been developed and are being used successfully. Cross-discipline and cross-Node searches were mentioned as particular areas in need of improvement. More on this topic will be said below under the Milestones.

2 - Each NASA planetary mission plan end-to-end archive generation and achieve on-time delivery of documented and validated PDS-compliant archives.

The PDS Nodes were very positive with regard to this goal, generally regarding it as having been accomplished.

3 - Generation, validation, and delivery of PDS-compliant derived products by NASA mission instrument teams and PI's from data analysis programs as a standard practice.

The PDS Nodes had mixed opinions regarding this goal. Most felt that good progress has been made, especially in the past few years as NASA AOs have increasingly emphasized archiving. However, it was noted that validation is still mostly done by the Nodes, which may indicate room for improvement with regard to validation tools for the use of data providers.

4 - Establish and maintain coordination with domestic and international counterparts to assure interoperability making available data from all planetary missions (i.e., NASA, ESA, and other agencies).

The PDS Nodes felt that this goal is well in hand, citing significant participation in international data archiving bodies such as IPDA (**see Finding X "Integration with Other Archives").**

Milestone for Data Access and Distribution

1 - PDS will support a reduced set of file structures/formats for science data, with the selection based on community input.

The current PDS had mixed opinions regarding this milestone, though good progress was acknowledged. On the positive side, PDS4 is explicitly designed to do this, although some Nodes emphasized that careful stewardship is required to prevent an unnecessary and detrimental proliferation of file formats, while other Nodes stated that community input has not been a major consideration in PDS decisions regarding file formats. **(See Finding IV "PDS File Formats and Translation Software").**

2 - All Discipline Node catalogs will provide search capabilities across all their holdings at a fine level of granularity, including geometric constraints.

The PDS Nodes agreed that some progress has been made with regard to this milestone but that more remains to be done. A major factor that limits cross-discipline and even cross-instrument search is that, in many cases, data providers were not required by NASA to generate metadata that would enable effective search. Some Nodes have addressed this situation by generating metadata themselves, so that it can be consistent across data sets, but this is a major undertaking that requires careful customization for each data-providing instrument. **(See Finding IX "Data Discoverability")**

3 - Basic data visualization tools will be integrated into the search process, so that most data products can be evaluated quickly by the user before selection for delivery. The PDS Nodes reported variable progress. Some have developed search tools that meet this milestone, while others have not made it a priority. **(See Section 3.2 "Data Discoverability")**

4 - Automatic translation tools will allow users to receive data in whatever suitable format they prefer.

The PDS Nodes generally agreed that this milestone has not been met, and some objected that the milestone was too expansively worded. Some translation tools are available, and in other cases translation is done by the Node and the resulting file formats are made available, but overall support is limited. (**See Finding IV "PDS File Formats and Translation Software")**

5 - Users will have the option of receiving data requests in the form of fully PDS-compliant custom volumes generated on-the-fly.

Although some PDS Nodes reported limited progress with regard to this milestone, there was a general feeling that this objective has been overtaken by events (e.g. by introduction of PDS4 and changes in storage and information transfer technology).

Milestone for Team Interfaces and Archiving

1- Mission archiving with PDS will be routine. Well established milestones that evolve into place based on PDS and community experience and inputs. Discipline Nodes will provide tools to the teams for generating the necessary metadata relevant to their discipline. A complete set of automated tools will be available to validate the metadata against PDS standards. The PDS Nodes were generally positive with regard to this milestone. However, many felt that tools for data providers remain deficient, though some of the tasks that would be done by these tools are instead done in-house by the Nodes. (**See Section 3.3 "Data Usability" and Findings I "Managing Expectations of PDS Usability", IV "PDS File Formats and Translation Software, and VIII "Documentation and Training"**)

2 - AOs need to require all instrument/mission teams to deliver fully calibrated, geometrically corrected data throughout the mission and to resubmit the data when they are improved, and to collaborate with the PDS on all data calibration/processing/retrieval systems designed for their team members. Data systems used by team members must seamlessly transition over to the PDS as the team's data sets become publicly available.

The PDS Nodes were very positive with regard to this milestone, though several noted that it is NASA rather than PDS that is responsible for many of the relevant actions. One Node pointed out that the milestone's language might be interpreted as requiring PDS to archive a mission team's pipeline, including software, which is generally regarded as outside the purview and capabilities of PDS (**See Finding XV "Archiving Software"**).

Milestone for Technical Infrastructure

1 - PDS architecture will be entirely distributed. Facilities at each Discipline Node will be capable of handling cross-discipline queries.

Some PDS Nodes felt that this milestone is being met or partially met, while others felt that it is not being met at all. **(See Section 3.2 "Data Discoverability" and Finding II "Access to Data" and Finding IX "Data Discoverability")**

2 - All data holdings will reside at a minimum of two different, geographically separated locations. Popular data sets will be continuously on line at two or more locations.

The PDS Nodes unanimously agreed that all data resides at multiple locations. However, several Nodes pointed out that online duplication or live mirroring does not make sense and is not done.

3 - All PDS software will be written in a manner that minimizes, and preferably eliminates, hardware and OS dependencies.

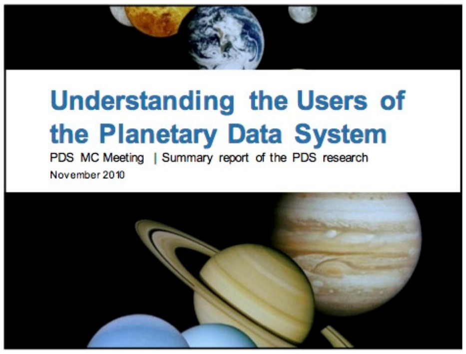
Some PDS Nodes felt that this milestone is being met or partially met, while others felt that it is not being met**.**

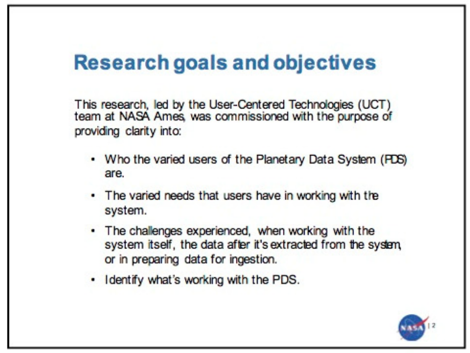
4 - PDS will upgrade its network connections regularly to provide fast delivery of data to users.

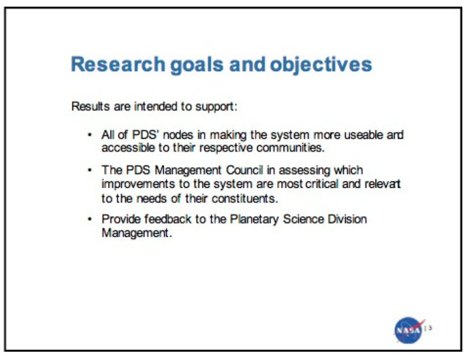
The PDS Nodes unanimously reported satisfaction with the speed at which they deliver data to their users.

5 - PDS will collaborate with other NASA entities to better assess and utilize existing and emerging storage technologies.

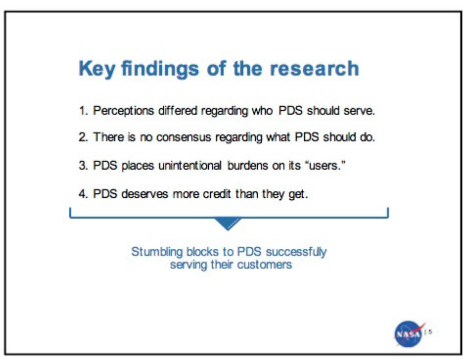
The PDS Nodes generally felt that this milestone is being met. **(See Finding XIII "Information Technology")**

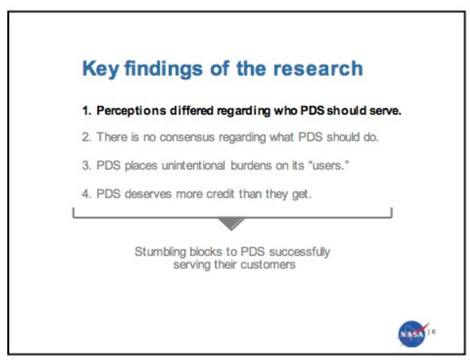

Understanding the Users of the Planetary Data System
PDS MC Meeting | Summary report of the PDS research
November 2010

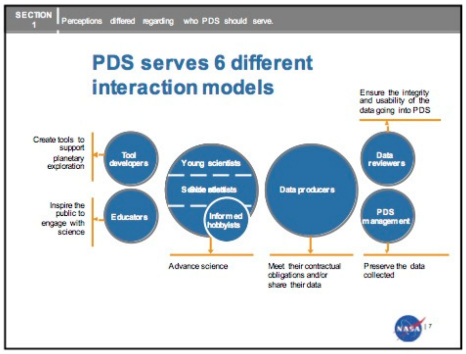

Research goals and objectives
This research, led by the User-Centered Technologies (UCT) team at NASA Ames, was commissioned with the purpose of providing clarity into:
• Who the varied users of the Planetary Data System (PDS) are.
• The varied needs that users have in working with the system.
• The challenges experienced, when working with the system itself, the data after it's extracted from the system, or in preparing data for ingestion.
• Identify what's working with the PDS.
2

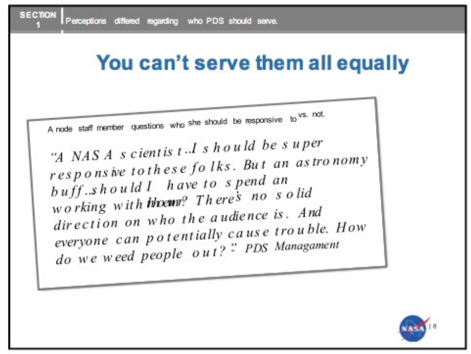

Research goals and objectives
Results are intended to support:
• All of PDS' nodes in making the system more useable and accessible to their respective communities.
• The PDS Management Council in assessing which improvements to the system are most critical and relevant to the needs of their constituents.
• Provide feedback to the Planetary Science Division Management.
3

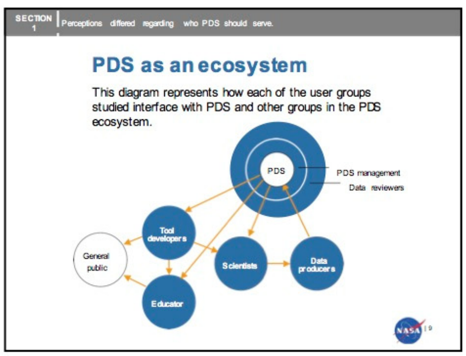

Research participants
53 people participated in the qualitative portion of this research.
81 people participated in the quantitative research.
Collage represents the institutions with which those individuals are affiliated.
4

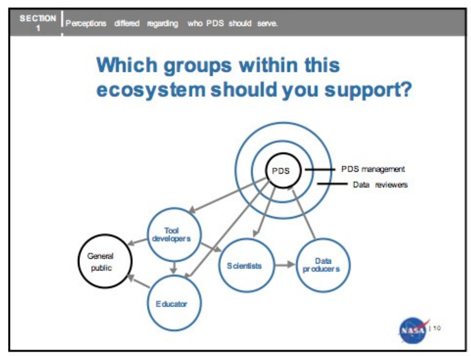

Key findings of the research
1. Perceptions differed regarding who PDS should serve.
2. There is no consensus regarding what PDS should do.
3. PDS places unintentional burdens on its "users."
4. PDS deserves more credit than they get.
Stumbling blocks to PDS successfully serving their customers
5

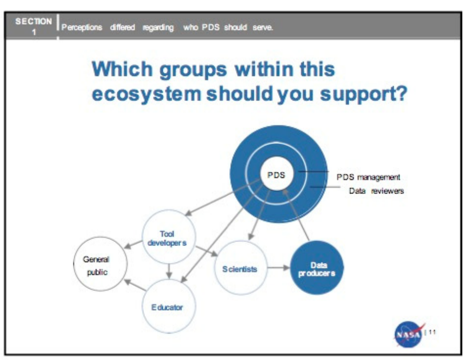

Key findings of the research
1. Perceptions differed regarding who PDS should serve.
2. There is no consensus regarding what PDS should do.
3. PDS places unintentional burdens on its "users."
4. PDS deserves more credit than they get.
Stumbling blocks to PDS successfully serving their customers
6

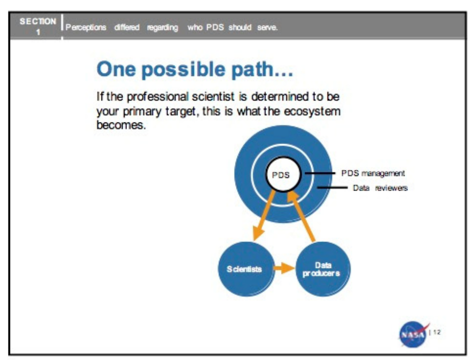

SECTION 1
Perceptions differed regarding who PDS should serve.
PDS serves 6 different interaction models
Ensure the integrity and usability of the data going into PDS
Create tools to support planetary exploration
Tool developers
Young scientists
Informed hobbyists
Data producers
Data reviewers
PDS management
Inspire the public to engage with science
Educators
Advance science
Meet their contractual obligations and/or share their data
Preserve the data collected
7


SECTION 1
Perceptions differed regarding who PDS should serve.
You can't serve them all equally
A node staff member questions who she should be responsive to vs. not.
"A NASA scientist..I should be super responsive to these folks. But an astronomy buff..should I have to spend an working with There's no solid direction on who the audience is. And everyone can potentially cause trouble. How do we weed people out?" PDS Managament
8


SECTION 1
Perceptions differed regarding who PDS should serve.
PDS as an ecosystem
This diagram represents how each of the user groups studied interface with PDS and other groups in the PDS ecosystem.
PDS
PDS management
Data reviewers
Tool developers
General public
Scientists
Data producers
Educator
9


SECTION 1
Perceptions differed regarding who PDS should serve.
Which groups within this ecosystem should you support?
PDS
PDS management
Data reviewers
Tool developers
General public
Scientists
Data producers
Educator
10

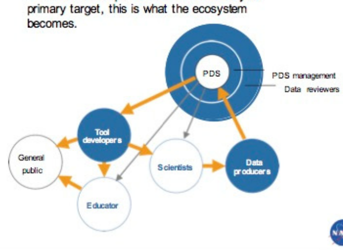

SECTION 1
Perceptions differed regarding who PDS should serve.
Which groups within this ecosystem should you support?
PDS
PDS management
Data reviewers
Tool developers
General public
Scientists
Data producers
Educator
11


SECTION 1
Perceptions differed regarding who PDS should serve.
One possible path...
If the professional scientist is determined to be your primary target, this is what the ecosystem becomes.
PDS
PDS management
Data reviewers
Scientists
Data producers
12

SECTION 1 | Perceptions differed regarding who PDS should serve.

Another possible path…

If the tool developer is determined to be your primary target, this is what the ecosystem becomes.



13

Key findings of the research

1. Perceptions differed regarding who PDS should serve.
2. **There is no consensus regarding what PDS should do.**
3. PDS places unintentional burdens on its "users."
4. PDS deserves more credit than they get.

Stumbling blocks to PDS successfully serving their customers

14

SECTION 2 | There is no consensus regarding what PDS should do.

To answer what PDS should do, you first have to ask…

1. What is PDS for people today?
2. What is doable for PDS today?

15

SECTION 2 | There is no consensus regarding what PDS should do.

What is PDS?

16

SECTION 2 | There is no consensus regarding what PDS should do.

What is PDS?

"It's the authoritative data archive for me. It's an awesome long-term archive, [but I liken it to a library] full of books with no spines on the books." - Scientist

17

SECTION 2 | There is no consensus regarding what PDS should do.

What is PDS?

"I've always perceived PDS to be holders of data. I never thought of PDS as enabling research." - Tool developer

18

What is PDS?

"I see it as a national archive. All the data is there, but stored on punch cards." Data provider

What is doable for PDS?

What is doable for PDS?

"I think it's PDS' responsibility to provide data in a format that can be transformed, but not their responsibility to provide the tools themselves."
- PDS Management

What is doable for PDS?

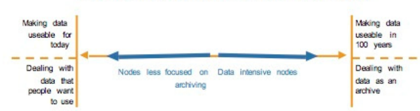
"[We] struggle to just do a good job archiving. Nodes don't have the resources for making derived products." - PDS Management

What is doable for PDS?

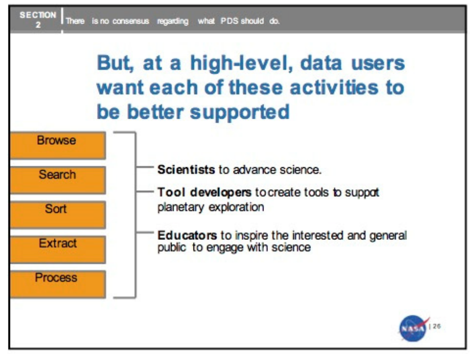
"Early on in PDS, PDS was mandated to be strictly about archive. We have basic interfaces to get at the data. [Customer] expectation now is that the interface will be much more advanced. Funding isn't aligned to deliver these interfaces."
- PDS Management

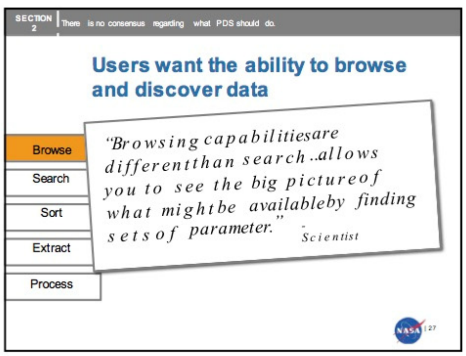
What is doable for PDS?

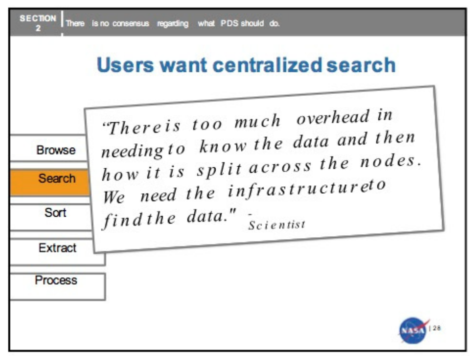
It became increasingly clear that PDS users need data to be useable and easily accessible. However, perceptions differed on where PDS should be, and how it can realistically play, on the data useable for today vs. in a 100 years continuum.

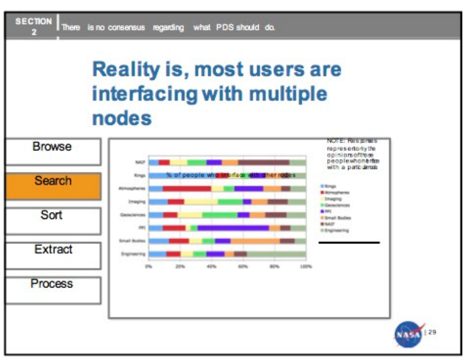


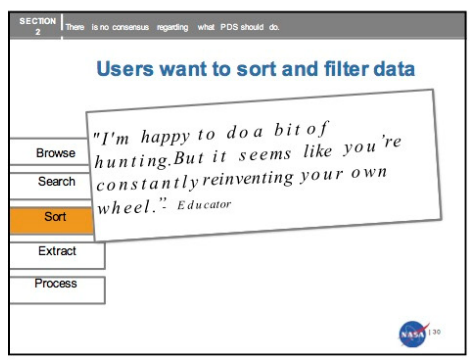

SECTION 2
There is no consensus regarding what PDS should do.
But what does it *really* mean when a data user says, "I want more useable and accessible data?"
It means something more than improved search…
25


SECTION 2
There is no consensus regarding what PDS should do.
But, at a high-level, data users want each of these activities to be better supported
Browse
Search
Sort
Extract
Process
Scientists to advance science.
Tool developers to create tools to support planetary exploration
Educators to inspire the interested and general public to engage with science
26


SECTION 2
There is no consensus regarding what PDS should do.
Users want the ability to browse and discover data
Browse
Search
Sort
Extract
Process
"Browsing capabilities are different than search..allows you to see the big picture of what might be available by finding sets of parameter." - Scientist
27


SECTION 2
There is no consensus regarding what PDS should do.
Users want centralized search
Browse
Search
Sort
Extract
Process
"There is too much overhead in needing to know the data and then how it is split across the nodes. We need the infrastructure to find the data." - Scientist
28


SECTION 2
There is no consensus regarding what PDS should do.
Reality is, most users are interfacing with multiple nodes
Browse
Search
Sort
Extract
Process
29


SECTION 2
There is no consensus regarding what PDS should do.
Users want to sort and filter data
Browse
Search
Sort
Extract
Process
"I'm happy to do a bit of hunting. But it seems like you're constantly reinventing your own wheel." - Educator
30

SECTION 2 | There is no consensus regarding what PDS should do.

Users want more flexibility and efficiency when extracting data

Browse
Search
Sort
Extract
Process

"People want data in a format that their program will accept. I wish that you could select, within reason, the formats that you want." - Scientist

| 31

SECTION 2 | There is no consensus regarding what PDS should do.

Users want more guidance on how to process the data

Browse
Search
Sort
Extract
Process

"80% of data in PDS isn't sufficiently calibrated. It's not likely that you can do something with it if you're not on the team. You have to be a genius." - Educator

| 32

SECTION 2 | There is no consensus regarding what PDS should do.

And when they can't meet these needs with PDS, they engage workarounds

| 33

SECTION 2 | There is no consensus regarding what PDS should do.

And when they can't meet these needs with PDS, they engage workarounds

Workarounds	Browse	Search	Sort	Extract	Process
Use mission and third party websites and tools *"I use HRSC team's tools to get a preview and see what I want. I then download it through PDS." - Scientist* NOTE: Question in the mind of many is: What happens to these tools when the mission is over? Will PDS take over these tools?	●	●	●		

| 34

SECTION 2 | There is no consensus regarding what PDS should do.

And when they can't meet these needs with PDS, they engage workarounds

Workarounds	Browse	Search	Sort	Extract	Process
Go directly to the data volumes and create mini-PDS' on their local hard drives *One scientist shared his process: He cannot browse images on PDS, hence he almost creates another PDS for data sets of his particular interest. He has created a database of Cassini imagery. He loads his own meta-data into the database, so that one day when he wants to see all the images of Saturn's moon he can. His database makes it easier for him to search the data.*	●	●	●		

| 35

SECTION 2 | There is no consensus regarding what PDS should do.

And when they can't meet these needs with PDS, they engage workarounds

Workarounds	Browse	Search	Sort	Extract	Process
Use PDS tools not designed for them *We met a few geologists using Rings' Opus tool. One of them said, "I want latitude and longitude parameters. And it's slower than I would like, but sill better than my other options."*		●	●	●	

| 36

And when they can't meet these needs with PDS, they engage workarounds

Workarounds	Browse	Search	Sort	Extract	Process
Go directly to the node staff *From a conversation with the PPI node: "We call this the 'The Old Steve, Ask Steve' method. A conversation with Steve is almost needed to help people zero in on the data that would be useful to them. ...And it should be noted that it's not just students using the 'Old Steve' method. Lots of senior people are using the 'Steve system' too."*		●	●	●	

And when they can't meet these needs with PDS, they engage workarounds

Workarounds	Browse	Search	Sort	Extract	Process
Bypass PDS all together and go directly to the mission teams *NOTE: Many of the more senior scientists commented on doing this, but as one so eloquently said, "Problem is ...helping people interpret data in PDS is not the responsibility of these folks [the mission team]. Many say, 'This is not part of my job description.' They want to know 'what's in it for me?' I shouldn't have to bribe people to use the data [that NASA has already paid for.]"*		●	●	●	●

And when they can't meet these needs with PDS, they engage workarounds

Workarounds	Browse	Search	Sort	Extract	Process
Give up on their science quest all together *NOTE: This was mentioned as a popular method in relation to grant proposals.*					

Now that's not to say that PDS isn't doing some great things…

"Going back to the early days of PDS would not be a good thing." For as much as PDS has room for improvement, where it is today is far better than where it came from." - Data producer

PDS already makes some data more useable and accessible

Tools mentioned unsolicited, were:

- Rings' Opus tool

"Enables interaction with the data."

"I like that I can sort by color."

"Love the running tab of results."

PDS already makes some data more useable and accessible

Tools mentioned unsolicited, were:

- Rings' Opus tool
- Ames' Volume Validator tool

"I discovered it [the validation tool] 2 months ago. It has changed my life completely. I get an immediate response to my questions and can quickly gain clarity on the standards."

PDS already makes some data more useable and accessible

Tools mentioned unsolicited, were:

- Rings' Opus tool
- Ames' Volume Validator tool
- Geoscience's Mars Orbital Data Explorer

"Have used it for footprints and have found it useful for browsing"

"Like that it begins to centralize search"

PDS already makes some data more useable and accessible

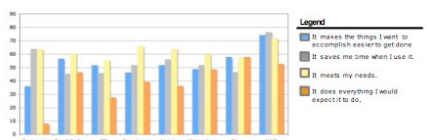
Tools mentioned unsolicited, were:

- Rings' Opus tool
- Ames' Volume Validator tool
- Geoscience's Mars Orbital Data Explorer
- Any sites/tools with Wget

"Best way when you're looking for 100s of images. Used to grab the cumulative index, drop it into a database and then filter. Now don't need to go to the cumulative index tab."

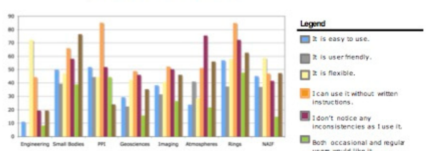
But here's some other data to help you make decisions regarding where improvements at your particular node can be directed…

Data users' perspective on PDS' usefulness



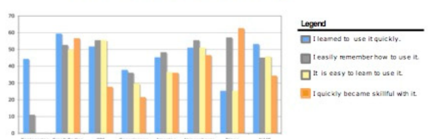
NOTE: Responses represent the aggregate of those provided by scientists, tool developers and educators (equaling a total of 76 of the 81 respondents).

Data users' perspective on PDS' ease of use



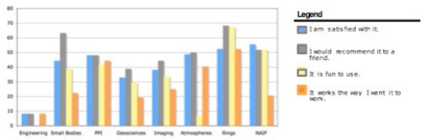
NOTE: Responses represent the aggregate of those provided by scientists, tool developers and educators (equaling a total of 76 of the 81 respondents).

Data users' perspective on PDS' ease of learning



NOTE: Responses represent the aggregate of those provided by scientists, tool developers and educators (equaling a total of 76 of the 81 respondents).

Data users' general satisfaction with PDS



NOTE: Responses represent the aggregate of those provided by scientists, tool developers and educators (equaling a total of 76 of the 81 respondents).

Key findings of the research

1. Perceptions differed regarding who PDS should serve.
2. There is no consensus regarding what PDS should do.
3. **PDS places unintentional burdens on its "users."**
4. PDS deserves more credit than they get.

Stumbling blocks to PDS successfully serving their customers

Data users feel burdened because…

It's not clear how and where to find the data.

Some data is duplicated across nodes and others not.

Some data is in PDS and others not.

Users want to search across data sets.

The node structure is too visible. Users want central search, without losing the benefits of on-call, node-specific expertise.

Users want to be certain that they've found all the data that there is to find.

Users want more meaningful categorizations for the data that does exist.

Data providers feel burdened because…

They are forced to work with multiple nodes if the mission demands it.

Process prevents them from putting the best quality data into PDS.

The accepted standards are broad and interpreted differently across the nodes.

Submitting data to PDS always requires a separate process.

There is no way to catalogue interesting and valuable data sets that don't discretely fit into the PDS standard.

The perception is that to add anything into PDS is a laborious process.

PDS Management feels burdened because…

PDS is in the precarious position of having responsibility without authority.

They question the sustainability of their structure.

They feel the pressure to develop things from scratch and all by themselves.

They consider their PDS node unique and do not actively try to learn from the successes of the other nodes.

They consider PDS unique, and thereby discount the value that could be gained from the successes of other science-based systems and fields.

Key findings of the research

1. Perceptions differed regarding who PDS should serve.
2. There is no consensus regarding what PDS should do.
3. PDS places unintentional burdens on its "users."
4. **PDS deserves more credit than they get.**

Stumbling blocks to PDS successfully serving their customers

There is no concerted effort to cross promote the successes of each node

The efforts that PDS makes to make data more useable and accessible are largely going unnoticed by the community because PDS doesn't announce what they are doing.

"PDS should make announcements about useful tools available to the community, not just new data sets available."

"I found the Mars Orbital Data Explorer on the Geosciences node. Have used it for footprints and found it useful for browsing. But came upon this by chance. Would be nice if these things were announced. Eventually found this through the main website."

When asked about where he would expect to learn about new tools: "Would expect PEN, maybe a booth at a conference, on mission websites, or cross-promoted on other sites of PDS."

There is no concerted effort to cross promote the successes of each node

Advertising new tools in only one location, whether on the node's site or via the central node's site is not sufficient.

People tend to have the node that has helped them most bookmarked. Only if they can't find what they are looking for there, do they look elsewhere.

PDS is not taking credit deserved for being the data engine behind other quality tools

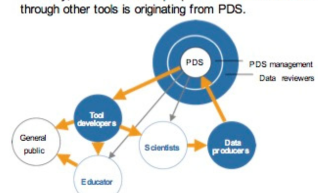
Tool developers admit that without PDS their tools wouldn't exist.

"If the PDS goes away, ASU's tools go away."

PDS needs to claim some of the credit deserved for making these possible – similar to Intel Inside, it's "PDS Inside."

How the "PDS inside" concept might play out...

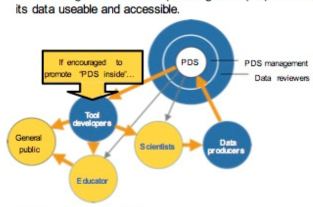
Currently, it's not visible to people that the data distributed through other tools is originating from PDS.



NOTE: Diagram represents the PDS ecosystem.

How the "PDS inside" concept might play out...

PDS could gain credit for empowering other people in making its data useable and accessible.



NOTE: Diagram represents the PDS ecosystem.

In conclusion, PDS needs to...

...gain alignment in who they serve.

...define explicitly what they are responsible for.

...reduce the unintentional burdens placed on their "users."

...get credit for the great things that they do.

These actions will enable PDS to more successfully serve their customers

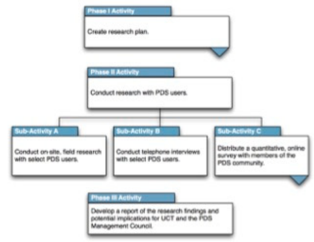
For further discussion about the content provided, contact Brianna Sylver or Jay Trimble.

Brianna Sylver
Sylver Consulting
brianna@sylverconsulting.com
Telephone: 312-239-0346

Jay Trimble
NASA Ames
jay.p.trimble@nasa.gov
Telephone: 650-604-6060

| 61

Research methodology engaged



| 62

Research participants

Took a broad view at defining who the potential "users" of the PDS are.

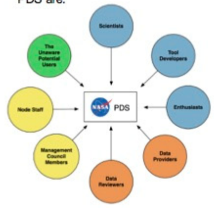


Diagram represents how the study was sampled. Color represents the various experiences we expected people to have with the system.

| 63

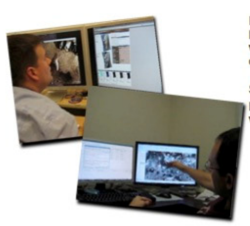
Qualitative research protocol

Explored topics such as:

- The intersection of PDS in the participant's workflow.
- Past experiences with PDS–both good and bad.
- Impression on who PDS' target user is.
- Thoughts on what's in and out of scope for PDS.
- Concerns related to the sustainability of the system over time.
- Perceptions around PDS node structure.

| 64

Qualitative research protocol



For many of those we met in person, we asked them to walk us through the tasks they do most frequently with PDS.

So, we saw the workarounds people engage to get their work done.

| 65

Quantitative research protocol

Leveraged a standardized, usability protocol.

Lund, A.M. (2001) Measuring Usability with the USE Questionnaire. STC Usability SIG Newsletter, 8:2

Protocol designed to assess the usefulness, satisfaction, and ease of use of PDS.

Has been successfully used in multiple domains.

Protocol certified by the UPA and IEEE.

| 66

See About[58] and Organization[59]

About the PDS

Overview

The Planetary Data System (PDS) is an archive of data products from NASA planetary missions, which is sponsored by NASA's Science Mission Directorate. We actively manage the archive to maximize its usefulness, and it has become a basic resource for scientists around the world.

All PDS-produced products are peer-reviewed, well-documented, and easily accessible via a system of online catalogs that are organized by planetary disciplines.

PDS technology has lowered both the cost and risk for large archives through online storage and tools. We use standards for describing and storing data that are designed to enable future scientists who are unfamiliar with the original experiments to analyze the data, using a variety of computer platforms, with no additional support. These standards (PDS Standards Reference and Planetary Science Data Dictionary) address the data structure, description contents, media design, and a set of terms.

The PDS is currently operating under a Charter (PDF) that was drafted in 2006, and adheres to the following set of high-level requirements that define the characteristics and features of the PDS operational system.

- PDS Level One, Two, and Three Requirements "April 2014" (.pdf)
- PDS Level One, Two, and Three Requirements "April 2014" (.doc)

Though PDS does not fund the production of archive data from active missions, we work closely with project teams to help them design well-engineered products that can be released quickly.

While most of our products can be ordered automatically, PDS provides teams of scientists to help users select and understand the data. We also offer special processing for products tailored to the needs of individual users.

Structure

PDS Project Management is assigned to the Solar System Exploration Data Services Office at the Goddard Space Flight Center.

PDS is a federation of 10 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 Facility Node. Additionally, PDS includes a special function supported by a Radio Science specialist.

Several of the nodes have "sub-nodes" to help with a specific aspect of the node's discipline. Several of the nodes oversee one or more data nodes, established for a short period of time to deliver a specific data collection to the PDS.

Each node is led by an expert in the subject discipline, and each has an advisory group made up of other practitioners of that discipline. Node selections are made every five years under a NASA Research Announcement.

One of the science node leaders is also assigned as the PDS project scientist; this position rotates every few years.

For more information about each node see PDS Organization.

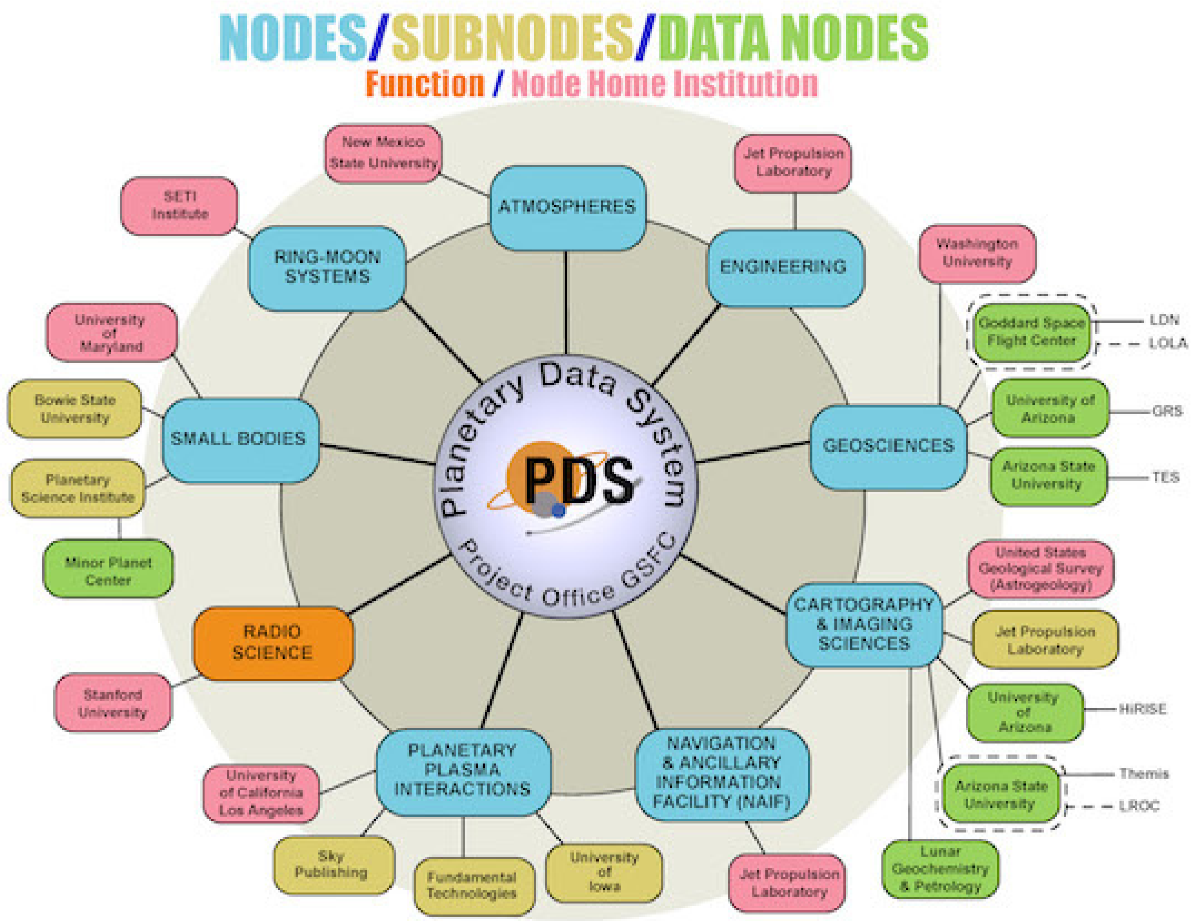
[58] https://pds.nasa.gov/about/about.shtml
[59] https://pds.nasa.gov/about/organization.shtml

PDS Organization

The Planetary Data System is organized as a federation of eight*Nodes*. In addition, *Subnodes* have been established to help the Nodes in some science discipline areas. And some of the Nodes oversee one or more *Data Nodes*, which are set up as a short-term facility to deliver specific data sets for archiving in the PDS.

Organizational Diagram

The following diagram shows the organization relationships among the Nodes, Subnodes, and Data Nodes, and the institutions at which they are located.



Details About the PDS Nodes

PDS Project Management	The PDS Project Management resides at Goddard Space Flight Center in the Solar System Exploration Data Services Office, Code 690.1

Engineering Node

The Planetary Data System (PDS) is a distributed information system with the Engineering Node located at the Jet Propulsion Laboratory in Pasadena, California. The Engineering Node provides systems engineering support to the entire PDS, handling global aspects such as standards (data, software, documentation, operating procedures), technology investigations, coordination and development of system-wide software, coordination of data ordering and distribution, catalog development and implementation, and maintenance of the PDS catalogs.

Atmospheres Node

The Planetary Atmospheres Node is responsible for the acquisition, preservation, and distribution of all non-imaging atmospheric data from all planetary missions (excluding Earth observations). The primary goal of the node is to make available to the research community the highest quality data possible. The data are reviewed and reformatted where necessary in order to meet the documentation and quality standards established by the PDS.

Geosciences Node

The Geosciences Node maintains data sets that are relevant to the geosciences discipline, the study of the surfaces and interiors of terrestrial planetary bodies. The Node works with planetary missions to help ensure that their geoscience data sets are properly documented and archived. The Node also restores and publishes selected data sets from past missions that are in danger of being lost. The Geosciences Node also provides information and expert assistance to researchers, and answers questions from the interested non-scientists as well. Derived image data, geophysics data, microwave data, spaceborne thermal data and spectroscopy data are archived at the lead node or at one of the subnodes.

Navigation and Ancillary Information Facility (NAIF) Node

The NAIF Node is responsible for design and implementation of the SPICE concept--a means for archiving, distributing and accessing observation geometry and related ancillary data used in mission design, mission evaluation, observation planning and science data analysis. Under PDS funding NAIF serves as the "ancillary data node"-- archiving and distributing the SPICE kernel files produced by numerous flight projects. SPICE kernel file distributions are accompanied by the latest NAIF Toolkit software. NAIF also produces and distributes generic ephemeris data sets for planets, satellites, comets and asteroids, using appropriate products of JPL's Solar System Dynamics Group as sources.

Cartography and Imaging Sciences Node

The Cartography and Imaging Sciences Node maintains and distributes the archives of planetary image data acquired from NASA's flight projects with the primary goal of enabling the science community to perform image processing and analysis on the data. The Node provides direct and easy access to the digital image archives through on-line remote-access tools by way of Internet services. The Cartography and Imaging Sciences Node provides digital image processing tools and the expertise and guidance necessary to understand and use the image collections. The Node is responsible for restoring data sets from past missions in danger of being lost. They also work with active flight projects to assist in the creation of their archive products and to ensure that their products and data catalogs become an integral part of the Cartography and Imaging Sciences Node's data collections.

Planetary Plasma Interactions Node

The Planetary Plasma Interactions (PPI) Node is responsible for acquisition, preservation, and distribution of fields and particle data from all planetary missions. The primary goal of the PPI Node is to make available to the research community the highest quality data possible. To insure the highest quality data, all data are reviewed and where necessary, reformatted to meet the quality standards established by the PDS.

Ring-Moon Systems Node

The Planetary Ring-Moon Systems Node is devoted to archiving and distributing scientific data sets relevant to planetary ring systems. The two major classes of ring data are images and occultation profiles, although a variety of additional data types (e.g. spectra, particle absorption signatures, etc.) are also of interest. A large fraction of our data sets are from the Voyager missions to the outer planets, but Earth-based and HST data sets are also represented. The Ring-Moon Systems Node also performs a variety of services to support research into these data sets. These services include developing on-line catalogs and information systems, filling orders for data, developing software tools, and coordinating special observing campaigns.

Small Bodies Node

The Small Bodies Node (SBN) is a distributed node that curates data sets and provides consulting expertise for comets, asteroids, and interplanetary dust. The Comet Subnode is located at the University of Maryland College Park. In addition to maintaining the combined archives of the SBN and supporting the SBN web site, the Comet subnode collects, formats, verifies and consults on datasets concerned with comet observations as well as providing support for active comet missions and observing campaigns. The Asteroid/Dust Subnode is located at the Planetary Science Institute in Tucson, Arizona. The Asteroid/Dust subnode collects, formats, verifies and reviews ground based and mission data pertaining to asteroids, transneptunians, small planetary satellites and interplanetary dust.

PDS REQUIREMENTS

1. PDS will provide expertise to guide and assist missions, programs, and individuals to organize and document digital data supporting NASA's goals in planetary science and solar system exploration.

1.1. Single Point of Contact: PDS will provide a single point of contact to each mission, program, agency, or individual (i.e., data providers) wishing to submit archival data

1.1.1. PDS will assign a lead node for each data provider submitting data to PDS

1.1.2. PDS will assign a lead individual, designated by the lead node, who is authorized to negotiate for PDS

1.1.3. The PDS lead node will delegate responsibility for subordinate contacts (e.g., instrument teams within a mission) to the appropriate PDS nodes

1.2. Expert Help: PDS will provide expert help in designing archival data sets

1.2.1. PDS will provide examples and suggestions on organization of data products, metadata, documentation and software

1.2.2. PDS will provide expertise in applying PDS standards

1.2.3. PDS will provide expertise to support the design of scientifically useful archival data sets

1.2.4. PDS will provide training to support the design of archival data sets for data providers on: PDS standards, tools and services

1.2.5. PDS will provide training to develop and maintain staff expertise in data engineering, standards and tools

1.3. Plans and Documents: PDS will assist data providers in developing archive plans, interface documents, validation procedures, and delivery schedules for PDS approval

1.3.1. PDS will provide examples of data management and archive plans (including interface documents, procedures, schedules and templates)

1.3.2. PDS will determine whether data management and archive plans and relevant interface documents meet PDS requirements

1.3.3. PDS will provide criteria for validating archival products

1.3.4. PDS will coordinate with the data providers to establish schedules for delivery of archival products to the PDS

1.3.5. PDS will coordinate with data providers to establish schedules for public release of archival products

1.4. Archiving Standards: PDS will have archiving standards for planetary science data

1.4.1. PDS will define a standard for organizing, formatting, and documenting planetary science data

1.4.2. PDS will maintain a dictionary of terms, values, and relationships for standardized description of planetary science data

1.4.3. PDS will define a standard grammar for describing planetary science data

1.4.4. PDS will establish minimum content requirements for a data set (primary and ancillary data)

1.4.5. PDS will, for each mission or other major data provider, produce a list of the minimum components required for archival data

1.4.6. PDS will develop, publish and implement a process for managing changes to the archive standards

1.4.7. PDS will keep abreast of new developments in archiving standards

[60] https://pds.nasa.gov/roadmap/pds_level123_requirements_20140410.pdf

1.5. Archiving Tools: PDS will have tools to assist data producers in assembling, validating, and submitting archival products

1.5.1. PDS will provide tools to assist data producers in generating PDS compliant products

1.5.2. PDS will provide tools to assist data producers in validating products against PDS standards

1.5.3. PDS will provide tools to assist data producers in submitting products to the PDS archive

1.5.4. PDS will provide documentation for installing, using, and interfacing with each tool

2. PDS will collect suitably organized and well-documented data into archives that are peer reviewed and maintained by members of the scientific community.

2.1. Solicit: PDS will seek complete and comprehensive archives from data providers consistent with interests and resources available.

2.1.1. PDS will compare proposed archival submissions against nominal content standards for similar archives and will seek augmentations when the submission is deficient

2.1.2. PDS will identify and maintain a list of proposed planetary science data sets to be added to the archive

2.1.3. PDS will work with relevant NASA program officials to ensure that products resulting from data analysis programs are submitted to the archive

2.1.4. PDS will provide a mechanism for the planetary science community to propose new additions to the archive

2.2. Receive: PDS will receive, acknowledge and track data submissions.

2.2.1. PDS will develop and publish the procedures for delivery of data to the PDS

2.2.2. PDS will track the status of data deliveries from data providers through the PDS to the deep archive

2.2.3. PDS will provide the necessary resources for accepting data deliveries

2.3. Validation: PDS will validate data submissions to ensure compliance with standards.

2.3.1. PDS will develop and publish procedures for determining syntactic and semantic compliance with its standards

2.3.2. PDS will implement procedures to validate all data submissions to ensure compliance with standards

2.4. Peer Review: PDS will conduct peer reviews of all submissions of archival data to ensure completeness, accuracy, and scientific usability of content.

2.4.1. PDS will develop and publish procedures for peer review of archival products (which includes all data submissions and ancillary information)

2.4.2. PDS will establish success criteria for peer review of archival products

2.4.3. PDS will implement peer reviews, coordinated and conducted by the lead node to ensure completeness, accuracy and scientific usability of content

2.4.4. PDS will publish a summary of the results of each peer review

2.4.5. PDS will track the status of each peer review

2.5. Acceptance: PDS will accept or reject submitted data.

2.5.1. PDS will develop and publish procedures for accepting archival data

2.5.2. PDS will implement procedures for accepting archival data

2.5.3. PDS will inform a data provider why a rejected archival product does not meet archiving standards

2.6. Catalog: PDS will maintain a catalog of accepted archival data sets.

2.6.1. PDS will develop and publish procedures for cataloging archival data

2.6.2. PDS will design and implement a catalog system for managing information about the holdings of the PDS

2.6.3. PDS will integrate the catalog with the system for tracking data throughout the PDS

- 2.7. Storage: PDS will provide appropriate storage for its archive.
  - 2.7.1. PDS will develop and publish procedures for storing archival data
  - 2.7.2. PDS will maintain appropriate storage for the PDS archive
  - 2.7.3. PDS will review its storage capacity and its anticipated storage requirements on a yearly basis
  - 2.7.4. PDS will maintain appropriate storage for non-archived data managed by the PDS
- 2.8. Architecture: PDS will maintain a distributed architecture based on scientific expertise
  - 2.8.1. PDS will maintain a distributed archive where holdings are maintained by Discipline Nodes, specializing in subsets of planetary science
  - 2.8.2. PDS will maintain a distributed catalog system which describes the holdings of the archive
  - 2.8.3. PDS will provide standard protocols for locating, moving, and utilizing data, metadata and computing resources across the distributed archive, among PDS nodes, to and from missions, and to and from the deep archive
  - 2.8.4. PDS will work with other space agencies to provide interoperability among planetary science archives
  - 2.8.5. PDS will provide an integrated on-line interface that provides information about and links to its data, services, and tools
  - 2.8.6. PDS will implement common and discipline-specific services within the distributed architecture
  - 2.8.7. The PDS architecture will enable non-PDS developed tools to access PDS holdings and services
  - 2.8.8. The PDS architecture will enable computational services on selected archival products
- 2.9. External Controls: PDS will adhere to applicable federal statutes, NASA policies and Memoranda of Understanding with other organizations.
  - 2.9.1. PDS will accept and distribute only those items which are not restricted by the International Traffic in Arms Regulations (ITAR)
  - 2.9.2. PDS will ensure that online interfaces comply with required NASA guidelines
  - 2.9.3. PDS will meet U.S. federal regulations for the preservation and management of data.
  - 2.9.4. PDS will fulfill obligations detailed in any applicable NASA Memorandum of Understanding (MOU)
- 2.10. System Development and Operations: PDS will follow best practices in system and software engineering for developing and operating the system
  - 2.10.1. PDS will monitor the system and ensure continuous operation
  - 2.10.2. PDS will identify and adopt technology standards (e.g., hardware and software) for the implementation and operations of the entire PDS system
  - 2.10.3. PDS will ensure that appropriate mechanisms are in place to prevent unauthorized users from compromising the integrity of PDS systems and data

3. PDS will make these data accessible to users seeking to achieve NASA's goals for exploration and science.

- 3.1. Search: PDS will allow and support searches of its archival holdings
  - 3.1.1. PDS will provide online interfaces allowing users to search the archive
  - 3.1.2. PDS will provide online interfaces for discipline-specific searching
  - 3.1.3. PDS will allow products identified within a search to be selected for retrieval
- 3.2. Retrieval: PDS will facilitate transfers of its data to users
  - 3.2.1. PDS will provide online mechanisms allowing users to download portions of the archive
  - 3.2.2. PDS will provide a mechanism for offline delivery of portions of the archive to users
  - 3.2.3. PDS will provide mechanisms to ensure that data have been transferred intact
- 3.3. Services: PDS will provide value added services to aid in using archive products.
  - 3.3.1. PDS will provide expert help in use of data from the archive

3.3.2. PDS will provide a capability for opening and inspecting the contents (*e.g.* label, objects, groups) of any PDS compliant archival product

3.3.3. PDS will provide tools for translating archival products between selected formats

3.3.4. PDS will provide tools for translating archival products between selected coordinate systems

3.3.5. PDS will provide tools for visualizing selected archival products

3.3.6. PDS will provide a mechanism for notifying subscribed users when a data set is released or updated

3.3.7. PDS will solicit input from the user community on services desired

4. PDS will ensure the long-term preservation of the data and maintain their usability.

4.1. Long-Term Preservation: PDS will determine requirements for and ensure long-term preservation of the data

4.1.1. PDS will define and maintain a set of quality, quantity, and continuity (QQC) requirements for ensuring long term preservation of the archive

4.1.2. PDS will develop and implement procedures for periodically ensuring the integrity of the data

4.1.3. PDS will develop and implement procedures for periodically refreshing the data by updating the underlying storage technology

4.1.4. PDS will develop and implement a disaster recovery plan for the archive

4.1.5. PDS will meet U.S. federal regulations for preservation and management of the data through its Memorandum of Understanding (MOU) with the National Space Science Data Center (NSSDC)

4.2. Long-Term Usability: PDS will establish long-term usability requirements and implement procedures for meeting them

4.2.1. PDS will define and maintain a set of usability requirements to ensure on-going utility of the data in the archive

4.2.2. PDS will develop and implement procedures for periodically monitoring the user community interests and practices and verifying the usability of the products in the archive

4.2.3. PDS will monitor the evolution of technology including physical media, storage, and software in an effort to keep the archiving technology decisions relevant within the PDS

4.2.4. PDS will provide a mechanism to upgrade products or data sets which do not meet usability requirements (*e.g.*, data sets from old missions)

L1/L2/L3: Approved by PDS Management Council
L1/L2: E-mail vote ending: 2005-07-06
L3: E-mail vote ending: 2006-05-26
L3 Updates: Approved by MC: 2006-08-03
L3 Updates: Approved by MC: 2010-03-26
Requirement 2.4 revision: Approved by MC 2014-04-10

As part of the responses to the CAN for competing the six science nodes in 2015, each proposal had to include an Executive Summary, outlining their plans for the future five years. The Engineering Node and NAIF were not competed but the Engineering Node did also provide such a Summary as part of its Senior Review. These were all completed prior to this Roadmap exercise but are attached here for reference.

The Atmospheres Discipline Node of NASA's Planetary Data System

Executive Summary

We propose to provide an organized, documented, and transparent archive of planetary mission data for the science community and entire world through the continued operation of NASA's Planetary Data System (PDS) Atmospheres Node at New Mexico State University. The Atmospheres Node will maximize NASA's return on its investment in planetary exploration and will aid in the expansion of knowledge of all communities. Our team is uniquely qualified to provide this service to the planetary science community due to

- Our broad understanding of and extensive background in the PDS,
- Our collective scientific expertise in the atmospheres of Venus, Mars, Titan, and the giant
- Planets,
- Our prominent role in the development of the latest archive standards (PDS4),
- The relationships we have cultivated with mission and instrument teams, and
- Our proposed initiatives for improving the processes for archiving higher order data
- Products and improving access to data by users.

The rationale for retaining a planetary atmospheres node of the PDS is based on several factors: a) the comparative nature of the research programs of many of planetary atmospheric scientists, b) the desire to maintain and enhance the ease of access to seemingly disparate data sets related to planetary atmospheres, and c) the shift within NASA's Planetary Science Division (PSD) to encourage interdisciplinary, comparative studies through a reorganization of Research and Analysis (R & A) programs based on crosscutting themes rather than target-based programs. Having NASA's mission data organized by science discipline enables the pursuit of comparative planetology science questions, which is consistent with NASA's own PSD science goals.

The data we will curate in the Atmospheres Node (ATMOS) consist of historical data already in the ATMOS archive, data from current missions that are actively archiving in ATMOS, and data from future missions with at least one instrument that will yield atmospheric science data. The historical and current mission data comprise an archive approximately 3.6 TB in size, arising from missions to every planet in the solar system and acquired over a time period of more than 40 years. The future data we will archive will likely double the size of the ATMOS holdings and will be provided by MAVEN, InSight, Mars2020, Cassini, Juno, the Europa Clipper, and atmospheric components of yet to be determined Discovery and/or

New Frontiers missions. In addition, we will archive a range of ground-based data acquired in support of NASA missions as well as a variety of derived or higher order data products resulting from NASA R & A programs.

The Atmospheres Node has played a critical role in the development and implementation of PDS4, and will continue to do so. As the PDS undergoes the current transition to the PDS4 archive standards, ATMOS has expanded its standard archiving efforts to include a) working with new missions as first use cases of the new PDS4 system, b) working with the PDS Engineering Node (EN) on the ongoing development and testing of new PDS4 software using ATMOS data as test cases, and c) leading the migration effort to exercise newly developed PDS4 software and tools, with the longer-term goal of migrating the majority of the ATMOS holdings to the PDS4 system.

The natural order of development for the ATMOS PDS4 effort involved first migrating test cases from PDS3 to PDS4 including all of our Mars Phoenix holdings and several accelerometer data sets. As we gained insight into how PDS3 and PDS4 worked from a migration standpoint, we were more prepared to lead development on the first PDS4 missions. We interfaced with LADEE as the lead node and assisted on the archive development for MAVEN. As a natural development progression, **as LADEE is ending and MAVEN prepares for its first delivery, ATMOS is leading the way with developing new, more comprehensive ways to disseminate PDS4 data to the science end-user**.

We have a comprehensive plan and approach for implementing and maintaining PDS archive standards that are state-of-the-art, compatible with those of other PDS nodes, and aligned with the international standards set forth by the International Planetary Data Alliance (IPDA). This includes a plan for moving forward with the implementation of the PDS4 standards with all the new missions that ATMOS deals with, as well as the migration of existing and new data sets that are submitted as PDS3 data sets. We will continue to work with other PDS nodes, most notably PPI and GEO, in developing the MAVEN, InSight and Mars 2020 archives and participating in mission reviews. This will assure that we are interpreting PDS4 standards in a similar manner to other nodes and will allow us to adopt desirable procedures that the other nodes develop.

Based on our extensive experience in working with missions and assessing best practices, we have developed a comprehensive set of guidelines that outline the roles and responsibilities for both the mission/instrument teams and the PDS Atmospheres Node. These guidelines provide a description of benchmarks such as when the data providers should establish a Data Archive Working Group, when they should provide the PDS with sample data files, when the PDS should provide the instrument teams with XML label templates, and when the PDS should conduct a peer review of the data. More importantly, we have developed these detailed guidelines to be interleaved with a general mission operations schedule so that the instrument teams, mission management, NASA Headquarters, and PDS personnel have a clear understanding of the deliverables and tasks at each stage of the archive development. We are adopting a parallel strategy for working with providers of derived data products, which will improve the yield of programs such as NASA's Planetary Data Archiving, Restoration and Tools (PDART) program.

We have developed a thorough and unbiased peer review plan for data submitted for archiving at ATMOS, with the goal of optimizing the quality of a data set and its documentation. We have outlined clear procedures for populating a review panel to ensure a high-quality review with no conflicts of interests. This group reviews the data based on explicit instructions from ATMOS, provides written comments to the

ATMOS review panel chair, and participates in a teleconference to define any liens and discuss the corrections and modifications of the data that are needed to ultimately lead to their certification.

Our archive focus is to fully implement the PDS4 capabilities while making data provision as efficient as possible and data access as transparent as possible. We will achieve this through the implementation of several innovative approaches:

- The Labeling And Registration Of Resource Product Pages,
- The Development Of Mission-Specific Help Pages,
- The Population Of Cross-Mission Matrices Focused On Science Goals, And
- The provision of multiple download options for a particular data set tailored to the interests and objectives of the user.

The Atmospheres Node has taken the lead on the exploitation of PDS4 capabilities to generate resource product labels. A unique philosophy to PDS4 is "Everything is a Product," and we employ this idea to its fullest potential. For the first time, PDS can allow mission and instrument help pages to be registered within the PDS4 system. These "Resource Products" allow not only the data products to be found via search, but also the web resources needed for finding more information about how to use the data effectively. This aspect of PDS4 is being developed with a joint effort between ATMOS and EN to effectively create resource product labels that efficiently register helpful web resources so that they can be seen in search results. With clever manipulation of the search result hierarchy at EN, we can assign our help pages and introductory web resources a high priority, ensuring that they appear at the top of the search results when presented to the end-user in addition to the most relevant requested data.

ATMOS is committed to developing data ingestion and access tools that will utilize the full capacity of PDS4 while serving as a source of assistance for our data suppliers and data users. We also will aggressively adapt PDS-wide software, evaluate tools posted on the tool registry of the IPDA, and direct our users via web page advice to current tools that are available outside the PDS that would facilitate access to a specific data set. Examples of tools that we will develop at ATMOS over the proposed award period include a tool to assist data providers in setting up their PDS4 bundles, a PDS4 archiving effort quantification tool, a documentation assessment tool, a PDS3-to-PDS4 translator tool, and the Easy Labeling System for Atmospheres.

The Atmospheres Node has been located at New Mexico State University since 1995 under the leadership of Dr. Reta Beebe. During the proposed award period Node leadership will transition to Dr. Nancy Chanover, who has been Beebe's Deputy PI since 2009. Chanover has been involved in the PDS as a data provider, data reviewer, and PDS data user for more than two decades. She will manage the staff of four research scientists, two faculty Co-Investigators, one graduate student, and three undergraduates to ensure optimal support for both the data providers and the users of the PDS Atmospheres Node. ATMOS also will make use of a seven-member Advisory Group, whose members have both expertise in a wide range of atmospheric science research areas and experience with the PDS. The Advisory Group members have agreed to provide feedback on specific aspects of the ongoing development of the archives, provide suggestions for peer review panels, address issues on web pages related to their areas of expertise, and take part in an annual assessment. The Atmospheres Node will prepare and transmit an annual status report and arrange for a teleconference with the Advisory Group to receive feedback and suggestions for improvement.

The PDS Atmospheres Node at NMSU has been and will remain committed to broadening the participation of underrepresented groups in NASA space science missions, research, and education. **NMSU is a designated accredited postsecondary minority institution by the U.S. Department of Education, with a Hispanic enrollment of 49% (2014-2015), and as such is the only minority serving institution with major representation within the PDS.** We have achieved a high participation and persistence among our underrepresented minority students and staff members by encouraging them to take advantage of upgrading their skill levels and working closely with them to assure they follow developments in planetary science. Undergraduate students are treated as professional employees and are supported to attend professional meetings, interact with potential users, and participate in working groups. By continuing to recruit and train high quality minority students at NMSU, ATMOS will contribute to developing greater diversity in the STEM skilled work force.

The Planetary Data System Geosciences Node at Washington University in St. Louis

Executive Summary

The Geosciences Node at Washington University in St. Louis has been an integral part of the PDS enterprise since NASA first assembled a data system to archive and distribute planetary data in response to recommendations from the National Research Council [Bernstein et al., 1982; Arvidson et al., 1986]. The Geosciences Node focuses on planning, validating, preserving, and making available archives that pertain to understanding the surfaces, interiors, and dynamics of Mercury, Venus, Earth's Moon, and Mars (pds-geosciences.wustl.edu). The node also collaborates with mission-funded data nodes at Arizona State University for Mars Global Surveyor Thermal Emission Spectrometer data, the University of Arizona for Odyssey gamma ray and neutron data, and Goddard Space Flight Center for Lunar Reconnaissance Orbiter Lunar Orbiter Laser Altimeter and radio science data. We also maintain close working relationships with the PDS Radio Science (RS) Advisor to ensure that we properly archive RS data. We currently have online archives totaling ~168 TB (terabytes) with users retrieving ~5 TB/month from our PDS Geosciences Node web site, and Orbital Data Explorer (ODE) and Analyst's Notebook (AN) interfaces. We currently work with eight active or recently completed missions, two developing missions (Mars Insight Lander and Mars 2020 Rover), 43 instrument teams, and 17 researchers who are restoring mission data sets or generating archives from Earth-based and laboratory observations.

We propose to continue the Geosciences Node, focusing on working with NASA missions to plan, receive, validate, and make available archives of use to the community using PDS4 standards. This work will include continued co-chairing with mission personnel and participating in Data and Archive Working Groups, including those for Opportunity and Curiosity rovers, Odyssey and Mars Reconnaissance Orbiters, Mars 2016 Insight Lander, the Mars 2020 Rover, Lunar Reconnaissance Orbiter, and depending on selections, a Discovery mission and a New Frontiers mission. We will also continue to coordinate European Space Agency's Mars Express archive deliveries to PDS. With the renewed emphasis on community recovery of older data sets and generation of new derived data sets, we will also continue and likely expand our interactions with researchers who successfully propose such efforts.

Orbital Data Explorer and Analyst's Notebook capabilities will be updated to include new archives and improved interfaces. We will also continue to develop Representational State Transfer (REST) capabilities that allow users to develop their own interfaces to our databases, e.g., an ArcMap interface at a user's institution. We will continue to access data at other nodes through both the ODE and AN interfaces so that users need not search multiple sites to find data.
We will continue as active participants in development and implementation of PDS4 standards. We will also continue participation in the International Planetary Data Alliance to promote access to data sets from international missions.

Management of the Geosciences Node will be under a senior planetary scientist (Arvidson) who has worked in archiving since the early 1980s. The Deputy Node Manager will be another senior planetary scientist (Guinness), augmented by an Operations Manager (Stein), an expert in advanced data processing techniques (O'Sullivan), and six staff members to cover planning, validation, curation, and making archives available. The work is enabled by an existing multimillion-dollar computation and data management system, with

backup facilities and IT (Information Technology) governance (security, data integrity, disaster recovery, and operations continuity) plans conformal with PDS requirements. We will continue to receive feedback from our advisory group, instrument teams, and researchers. This is done via face-to-face sessions, online forums and email, and attending meetings such as the Lunar and Planetary Science Conference.

Cartography and Imaging Sciences Node

Executive Summary

The objective of this new, 5-year program of work is to operate a science discipline node for the NASA Planetary Data System (PDS), specifically the "***Cartography and Imaging Sciences Node***" (or Imaging, IMG). Members of IMG have established expertise in archiving large volumes of data for the PDS and have been active participants in the forward-looking management of NASA's data archives for planetary exploration. Data for which IMG is responsible to PDS are being used continuously to support major scientific discoveries in planetary science. A few recent examples include identification and mapping of water on the "bone dry" Moon (M3, Pieters et al., 2009), recent activity in martian gullies (HiRISE, Dundas et al., 2012) indicating the frequent movement of volatiles at the surface of Mars, new impacts on the Moon and Mars (LROC, Robinson et al., 2015; HiRISE, Dundas et al., 2014) suggesting that there are active surface processes that must be monitored and understood in support of in-situ exploration of those bodies, and the recognition of relatively young lunar volcanic features (Braden et al., 2014), pointing to the recent shaping of the Moon by volcanic processes. Many of these results have completely changed our understanding of the origin and evolution of these planetary bodies.

Current IMG data holdings total ~825 TB and are conservatively projected to reach 1.3 PB in the next five years. IMG is dedicated to making these data and information resources accessible, discoverable, and usable by the public and thus it plays a major role in NASA's plan to improve access to data and publications resulting from NASA-funded scientific research (OSTP Memorandum, 2013; NASA, 2014). Continued availability of IMG data and services as part of the PDS federation holdings will fuel scientific discovery, innovation, and possibly entrepreneurship in the United States and among US partners abroad.

The Astrogeology Science Center of the United States Geological Survey (USGS, Dept. Interior; Flagstaff, AZ) and the Instrument Software and Science Data Systems Section of the Jet Propulsion Laboratory (JPL), California Institute of Technology, will continue their partnership as the ***Cartography and Imaging Sciences Node*** (IMG) of the PDS. The leadership of IMG resides at USGS, with Dr. Lisa Gaddis serving as Principal Investigator (PI). Ms. Susan LaVoie will serve as Co-Investigator and Institutional PI at JPL. IMG will continue to work with Arizona State University, Tempe, in support of the THEMIS Data Node (TDN; Dr. Phil Christensen, PI) and the LROC Data Node (LDN; Dr. Mark Robinson, PI), as well as the University of Arizona in Tucson, AZ to support the HiRISE Data Node (HDN; Dr. Alfred McEwen, PI). In 2015, IMG added support for a new Data Node (DN), the "Lunar Geochemistry and Petrology Data Node" (LGP) [LGPDN, Kirsten Lehnert, Columbia University (CU)]. The LDP project was recently funded by PDART program (as "MoonDB") to archive Apollo sample data, with an IMG partnership to integrate their data with geospatial products and deliver them to the PDS. In the event of decommissioning of any or all of the DNs, transition plans are in place for TDN and HDN (and under revision for LDN and development for LGPDN) to ensure that their data and basic data delivery services remain available for users.

New team members will be added to IMG to support the strong need for additional scientific and technical demands of the evolving PDS system architecture (called PDS4). The IMG team, comprised of experienced data users and data technologists, brings both enthusiasm and expertise in planetary science research and data technology to the PDS and its user community. This IMG team has substantial involvement with planetary science research and NASA spaceflight missions, and it has unique capabilities in cartographic, geodetic, and photogrammetric processing of planetary image data, software development and digital data analyses, and informatics infrastructure development and maintenance. Through an early pilot study for what was to become the PDS, IMG has provided data archive services and expertise to NASA for more than 25 years.

As demonstrated in this proposal, this IMG continuation leverages substantial holdings in scientific and technical expertise, ongoing research, facilities, hardware and software infrastructure, and outreach projects at the host institutions. IMG staff members understand the imaging data needs of the planetary science community now and into the future, and we have the experience and facilities to meet those needs and to help maintain the direction and momentum of the PDS as a service provider to the research community. We intend to provide seamless service to the PDS and its data providers and users through the efforts of dedicated staff and hardware in the primary areas of mission interface, data delivery and support services, and archiving support for new projects to modernize and archive historic planetary data products. Planned IMG activities in the FY16 to FY20 timeframe will build directly on existing relationships with data providers and with other PDS science and support nodes. The IMG team will contribute to the continuing evolution of the PDS as an effective virtual institute by providing online access to image data holdings for all Discipline Nodes.

The primary goal of IMG is to meet PDS requirements (***Table 1***; PDS Roadmap, 2006) and challenges by providing data and services that enable widespread use of planetary image data. In this proposal, we describe IMG activities (with focus on FY16 to FY20) that support the mission and vision of PDS to efficiently archive and make accessible data from and for NASA's planetary missions. These activities support NASA management, operations, and research and data analysis programs by facilitating release and enhancing access to data that inspire new planetary exploration, research and scientific discoveries.

The size of the growing IMG and PDS archive means that a major focus of future IMG work addresses the need for improved methods of finding just the right data products. Challenges include the need for new tools for efficient data searches and data mining of the vast amounts of PDS data and the science expertise needed to help PDS develop and maintain discipline-level data dictionaries and to help science users efficiently access desired data and archive the growing numbers of geospatial research products derived from PDS data. Significant contribution to the innovative development of solutions to "big data" archives in support of research in planetary science is expected from the efforts outlined. Development of these enhanced capabilities by a seasoned, dedicated team will provide a comprehensive system that will increase the ability of planetary scientists to identify and use planetary image data from past, present and future space missions.

The planned IMG activities are directly relevant to the PDS prime directives of archiving, distributing, and making available data from past and present NASA planetary space missions. IMG team members are involved with planetary flight projects during all phases of a mission investigation. IMG continues to restore data from historical planetary missions and to develop data labeling and formatting standards to improve archival accessibility and to reduce cost. IMG team members and their facilities sponsor workshops to train users in the use of access and analysis tools and to promote science research with PDS data. IMG provides support for planetary research by scientists at host facilities and at other institutions. IMG activities will be responsive to the needs of the planetary science community as they are represented by IMG scientists and members of a Science Advisory Committee. IMG expertise and services directly benefit users of data from past, present, and future NASA missions and help to ensure the long-term preservation and usability of planetary science data.

Table 1. PDS Requirements and IMG Activities.

PDS Level 1 Requirement	IMG Activities
PDS will provide expertise to guide and assist missions, programs, and individuals to organize and document digital data supporting NASA's goals in planetary science and solar system exploration.	• Provide a single POC for each instrument or proposal team • Provide expert help in designing and documenting archives (planetary and imaging science, archive standards) • Provide assistance in Archive Plans, I/F, Validation procedures and delivery schedules • Provide expertise in PDS3 and PDS4 archiving standards • Provide tools for producing and validating archive products, usable on a variety of platforms and systems
PDS will collect suitably organized and well-documented data into archives that are peer reviewed and maintained by members of the scientific community.	• Solicit complete and comprehensive archives from data providers • Validate submissions to ensure compliance with standards • Conduct and/or participate in peer reviews of data to ensure completeness, accuracy and scientific usability • Receive and track data submissions • Catalog data • Store data following best practices for integrity and security • Maintain a distributed architecture based on scientific expertise • Ensure adherence to Federal statutes and NASA policies • Follow best practices in system and software engineering for developing and operating the systems
PDS will make these data accessible to users seeking to achieve NASA's goals for exploration and science.	• Provide online interfaces allowing users to search the archive (Atlas) • Facilitate transfer of data to users thru bulk download, access to online repositories, standard interfaces to holdings (e.g., RESTful, GDAL) • Provide value-added services and expert assistance to aid in using archive products (MAP/MAP2, UPC, PILOT, POW, ANNEX, format and label transformation tools, ISIS, VICAR)
PDS will ensure the long-term preservation of the data and maintain their usability.	• Ensure long-term preservation of holdings thru data integrity checking, information security, backups (local, remote, deep), disaster recovery planning • Ensure long-term usability of holdings thru frequent interaction with the science community and monitoring of applicable technology evolution (h/w, s/w, storage, media)

Planetary Plasma Interactions (PPI) Node of the NASA Planetary Data System

Executive Summary

We herein propose to continue operation of the Planetary Plasma Interactions (PPI) Node of the NASA Planetary Data System for the next five years. The PPI Node archives data from NASA sponsored charged particles and electromagnetic fields instruments on NASA and international planetary spacecraft. Over the past 25 years the PPI Node has assembled a large archive of planetary data and it is still growing. Throughout its existence the PPI Node has been based on the concept of service to the scientific community:

Service to the NASA missions and investigators to help them create high quality archive products in an efficient and cost effective way.
Service to data users to help them access the data they need.
Service to future generations by archiving the data under standards that capture the important information about the data and preserve it.

This proposal describes the current PPI Node and how we plan to improve it over the next five years.

The Planetary Plasma Physics Discipline and its Archive

The PPI Node has assembled a large archive including data from 28 spacecraft at all eight planets, many of their moons as well as several comets, asteroids, and the interplanetary medium. There are 33 TB of PPI data comprising 632 data sets containing 2,792,489 products in 6,244,428 files. The data come from *in situ* charged particle and electromagnetic fields experiments. This data base supports the planetary plasma physics discipline. It fills a unique role within planetary science and within the Planetary Data System. The scientific discipline of planetary plasma physics is concerned with the physical processes in planetary magnetospheres and ionospheres. Time has proven that these data sets also are valuable to other disciplines in planetary science such as planetary interiors and neutral atmospheres. More recently planetary plasma physics has evolved into the study of comparative planetary magnetospheres and ionospheres. Now we are challenged to take our understanding of natural plasmas and extend them to the wide parameter range provided by the entire solar system. That requires the best quality data from the wide range of particles and fields instruments from throughout the solar system.

The PPI Node is currently archiving data from seven active missions including the MAVEN mission to Mars. PPI is the lead node for MAVEN and is coordinating the review and public release of the first archived data. In the immediate future PPI will provide support for the InSight mission scheduled for launch in March 2016. The newly selected Europa Clipper mission has particles and fields experiments. Several possible future missions have proposed charged particles and fields instruments. In addition to data directly from missions we will validate and archive data from Planetary Division data analysis and preparation (PDAP) opportunities. In the next five years we will bring the first ground based data into PPI. This will come from the NASA Planetary Division's Jove radio astronomy data base. We also will begin to provide access to planetary plasma models and simulations.

Ingesting Data into PPI by using PDS Standards

Over the past four years the PDS project has been developing a new data standard, PDS4. This improved version of the PDS archival model is based on the eXtensible Markup Language (XML) that is the standard for modern information systems. We have been active in the development of PDS4. The first two missions to use PDS4 were the LADEE mission (Atmospheres Node lead) and MAVEN **(**PPI Node lead**)**. PDS4 was being developed as these missions were preparing their data. It was unreasonable to ask the mission to write metadata to a standard that was still in development. Therefore in keeping with our philosophy to help where we can, we decided that the MAVEN mission would provide the data in the form used in the science community and that PPI would create the metadata. This led to PPI developing an archive version of the Common Data Format (CDF-A) that is consistent with PDS4 standards. CDF is widely used by the international community and we are working with that community to use CDF-A as they move toward adopting PDS4. This has given us important experience with PDS4 which we will share with the other PDS nodes. We also have worked out a plan to interact with missions when PDS4 is mature and the missions can generate PDS4 metadata on their own. During the next five years all of PPI**'s**_ _PDS3 compliant archive will be migrated to a PDS4 compliant archive. An important development over the past decade has been the founding of the International Planetary Data Alliance (IPDA). IPDA involves many countries involved with planetary missions. PPI is actively involved with the IPDA.

PPI has developed an efficient system for supporting missions in the design and preparation of their archival products. We work with them in writing the Project Data Management and Archive Plan (PDMAP) and PDS required documentation (Software Interface Specification (SIS) and the Interface Control Document (ICD)). We serve on the Data Archive Working Group and work with the data providers during all steps in preparing a data archive.

We strongly support the concept of peer review of the data products. Over the years this process has uncovered many problems that after correction helped create improved data products. The peer review is a cooperative process involving the data provider, PDS and scientists not involved in the mission. When minor problems are found the PPI Node corrects them so that the data providers can concentrate on the more serious problems. It has worked well.

Accessing and Using Planetary Plasma Data

The existing PPI Node web interface is a professionally designed layout based on 20 years of feedback from users. The web interface supports two ways to access the data. One is a "click and discover" system under which a user can select a planet, a mission, or an instrument type and home in on the data desired. The second and recently more popular approach is to use a word search (Google like) to find the data. In keeping with our philosophy to aid users in finding data we will improve the search and download capabilities. During the next five years we will enhance the word search by using technology we developed for the Heliophysics Division called "smart" search. This is an adaptive word based system that makes homing in on the desired data faster. The current system only allows users to download data using the file structure provided by the data provider. The improved system will allow downloads specified by time. We also plan improved graphics capabilities to aid in selecting the data. Finally for users needing to download data in large amounts we will implement a new download system called "mimic" that greatly increases efficiency. It also has the capability to check the data integrity of the data downloaded. We will add a "related to" capability to collect associated data and group them together. For instance one frequently needs magnetic field observations to interpret particle data. When a user finds the particle data of interest the "related to" feature will put the magnetic field data just a click away. For several years PPI has support data analysis web pages tailored to NASA program announcements. Those pages show the data that have been certified for use in a given program. Finally at the request of data users we are in the process of implementing special web pages for each mission giving a convenient and comprehensive view of all mission related data and information.

The data provided by the investigators is not always in the form that users need for their research. Therefore we have developed a series translation tools to place the data in common formats (e.g. CSV, VOTables). In the next period we will add translations to formats useful for IDL, MATLab and Python.

Finally we will move to archive software. Historically it has been difficult to archive software which tends to be short lived. However, recently technology has become available to archive software. We will work with the community developing these tools to develop a software archiving capability.

The PPI Team and Management

Managing a data archive requires two types of expertise and experience. First you need scientists who actively do research with the data being archived. Second you need expertise and experience in information technology. Best practices have shown that science data systems are best managed by scientists who have a strong interest in the resulting data products. In addition, science data archives require professionals who understand the technical challenges in preserving data in a usable form for the long term. Since PDS also serves data to the science community the technical team need expertise in data access and delivery. Operating a data archive requires a team of individuals with complementary experience who work efficiently together. We have formed such a team who has successfully and effectively managed the current PPI node from its inception.

The PPI Node has two specialized subnodes at the University of Iowa and Fundamental Technology Inc. The two teams have worked with PPI for years. They work well with the scientists and information technologists at UCLA. Each subnode has specific tasks and responsibilities. The management of these interactions is the responsibility of the PPI Node Manager. He has responsibility under the UCLA contract and grant management system to oversee the work. Payments are not made until he approves the data products being generated. In addition to checks during the year the PPI Node holds an annual meeting in San Francisco at the time of the AGU meeting at which the progress for the year is reported.

The PPI Node fully participates in PDS management activities. We submit monthly reports on data ingestion progress and monthly reports on system usage. We submit the annual report and give reports at PDS management council meetings.

Ring-Moon Systems Node

Executive Summary

We propose to continue serving as a Discipline Node of the Planetary Data System. Our plans build upon our long experience in addressing the needs of diverse users. Among these plans is the continued enhancement of OPUS, our signature "Outer Planets Unified Search" engine, which addresses the challenging problem of how to search for data products across a disparate collection of missions and instruments.

Our primary goals for the next five years are to:

- Continue improving the capability of PDS users to search for planetary data, with particular emphasis on new ways to describe data products and their relationships.
- Generate ancillary products that enable scientists to focus on research questions rather than on data processing questions.
- Bring color and motion to the archive through the new concept of "composite products."
- Enhance the scientific legacies of the Cassini and New Horizons missions.
- Expand the support that PDS provides for the planetary data sets from the Hubble Space Telescope and the James Webb Space Telescope.
- Preserve and simplify access to legacy data from NASA's earlier missions to the outer planets, including Pioneer, Voyager, and Galileo.
- Simplify the efforts of Earth-based observers, and the creators of small, derived data sets, to preserve their digital creations for posterity.
- Integrate all of our data sets and facilities into the new and powerful PDS4 system, and thereby make them available across all of the PDS Discipline Nodes.
- Streamline the process of migrating older PDS3 data sets into PDS4.
- Present compelling web content within an engaging and dynamic user interface.
- Enable our web services to interoperate seamlessly with other tools that PDS Nodes, scientists or members of the public might devise.
- Disseminate powerful, open-source software to streamline data archiving and to support data analysis.

To these challenges we will apply our extensive experience in building state-of-the-art software tools and data pipelines, in understanding the workings of instruments and missions, and in conducting our own scientific research using the same data sets that we curate. Having been key players in the design of the new PDS4 standards, we look forward to leveraging the capabilities of the new system.

Node Scope and Rationale

We are proposing an expanded scope that more accurately reflects the role that the Rings Node has long played within the organization. By adopting the name "Ring-Moon Systems Node," we will be better able to emphasize our support not just for studies of ring systems, but also for studies of any planetary system composed of multiple interacting bodies, small or large. From the viewpoint of the user community, the new name provides continuity with our past, but also emphasizes our extensive support for studies of planetary satellites in addition to rings. Our expanded emphasis on the dynamics and observing geometry of moons and rings continues to complement the existing PDS Node structure.

As a component of our expanded scope, we are planning a major initiative to preserve the finest Earth-based data about Uranus and Neptune. These "ice giants" are the focus of some of the most compelling research questions in planetary science. The planets share many traits in common, but differ in fundamental ways from Jupiter and Saturn. Because a major spacecraft mission to Uranus or Neptune is appearing to be more likely, we are planning to redouble our efforts to preserve the key data sets that will provide the historic context for a future mission to Uranus or Neptune.

Metadata Initiative

Our expansion of scope is driven in part by the recognition that data sets relevant to rings are almost always applicable to moons and planets as well. Ring photometry, small satellite astrometry, and feature tracking on a planet are commonly obtained from exactly the same products. The Rings Node has long recognized this overlap. It is one of the reasons that our search engine, OPUS, is PDS's primary search engine for Cassini's remote sensing observations. When a mission encompasses as many different targets and disciplines as Cassini does, it makes sense for one Node to take the lead on the search function. OPUS can find all the observations of a particular crater on Mimas, or all the observations of Saturn's north polar hexagon, just as easily as it can find the images of a particular ringlet.

OPUS's powerful search capabilities did not come easily. Most teams provide only minimal metadata to describe a product. For the Cassini data, we tackled this issue by developing powerful tools, built upon SPICE, that fully sample the field of view of each data product. As a result, OPUS provides much more comprehensive search results than would otherwise be possible.

We in the PDS have long recognized the gap between what our data providers deliver and what our users expect. One overarching goal for the next five years is to close this gap. This task has two facets: (1) to fill in critical missing metadata in old and current data sets, and (2) to open-source the tools so that future instrument teams can close this gap on their own.

Data Sets

Two current missions are of greatest interest to our user community: **Cassini** and **New Horizons**. In both cases, other Nodes have lead responsibilities but we collaborate closely and share the workload. Cassini is in its eleventh year touring the Saturn system, and it continues to send back exquisite data about the planet, rings and moons. The New Horizons spacecraft provided our most recent close-up data on the Jupiter system and will, of course, soon obtain unprecedented views of the Pluto system. At the time of launch, no one quite anticipated that Pluto and Charon lie at the center of a closely-packed system of four additional satellites, showing interesting phenomena including chaotic dynamics. For both of these missions, our primary role right now is in data validation and in developing the necessary search tools.

The same support remains needed for the data from several historic missions to the outer planets: **Voyager**, **Galileo**, and **Pioneers** 10 and 11. For each of these, our efforts will be devoted to ensuring the long-term usability of the data and metadata.

As PIs of a new PDART grant, we and our other Discipline Node collaborators are building the pipeline that will bring ~ 30,000 data products from the **Hubble Space Telescope** into the PDS. Just as this project completes, the **James Webb Space Telescope** (JWST) will launch. Because JWST and HST share the same data archive, we will be ready to incorporate the planetary products from JWST into PDS on Day 1.

We will also be archiving some key Earth-based data sets, including decades of Uranus and Neptune images from the W. M. Keck telescope, a large archive of Uranian ring occultation profiles, and 10 years of Uranus system astrometry. These will be valuable PDS archives in their own right, but they will also allow us to prototype methods for archiving smaller data sets efficiently.

PDS4 Migrations

PDS4 is the new standard for planetary data archiving; it is a key step in transforming the PDS into a modern, agile data archive. After making major contributions to its development, we have embraced the new standard and look forward to seeing it in full operation. To take advantage of the new capabilities, we will be migrating many important PDS3 data sets to PDS4.

Some PDS3 data sets, including Cassini ISS, Cassini VIMS and Voyager ISS, have been "orphaned" due to new restrictions on file formats. We have proposed, and are spearheading the effort to develop, an innovative solution to this problem: on-the-fly migration. This will enable these and other data files to be preserved in their original formats, avoiding costly duplication. In the proposed system, a user request will immediately trigger a conversion from the original PDS3 files to the new standard. This concept has been endorsed by the Engineering Node and the PDSMC.

Tools and Services

We have numerous plans to enhance our online services and tools. In continuing our commitment to providing diverse ways of accessing our work, we plan to release most of our tools as open source. We will also define and publish APIs (Application Programmer Interfaces) that enable anyone to have scriptable access to our tool or to embed our tools inside other web sites or interfaces of their own devising.

We will continue to upgrade OPUS with new data sets, new metadata, and new features, including an option to search for moving targets. We will also publish a complete API.

We will continue to enhance, and then open-source, our software library for performing sophisticated planetary geometry calculations.

We will define and implement a new class of "composite data products" which define particular kinds of relationships among other products. This will make it possible for users to search the archive for sets of products that represent movies, mosaics, color sets, etc.

We will upgrade our web site using tools that will make it easier to maintain, give it a fresh look, and make better use of screen real estate for screens large and small. We are also designing a "RESTful" interface which will make our pages and other on-line holdings available through a predictable API.

We will release online tools that enable users to enhance and examine our data products—including images and occultation profiles—interactively.

Perceived Significance

We bring a unique set of skills to the PDS. We are keenly focused on providing users with new ways to discover, access and analyze data. We are also seeking new ways to streamline the archiving process for data sets new and old. Not only will these contributions be of immediate benefit to PDS users and providers, but they can be spun off to benefit the other Discipline Nodes as well.

Small Bodies Node

Executive Summary

Introduction

The Small Bodies Node (SBN) of the PDS, managed by the University of Maryland and in collaboration with the Planetary Science Institute, has supported the archiving of spacecraft and ground-based observations of asteroids, comets, interplanetary dust, small satellites, dwarf planets, meteorites and related laboratory data for the last 20 years. We have pioneered the development of archiving tools and standards for the ingestion and dissemination of the most diverse collection of targets and phenomena in the PDS. We continue to reach out to top experts in small bodies science to ensure the data we archive will be accessible, usable and of value to the planetary community for decades into the future. We propose to continue this essential work, while taking advantage of new technologies and being responsive to evolving data collection methods and community needs.

Archiving Focus

The ensemble of small solar system bodies encompasses comets, asteroids, dwarf planets, TNOs, irregular satellites, and interplanetary and circumplanetary dust. These supposedly distinct classes of objects are not truly distinct, with fuzzy boundaries (e.g., comets and asteroids) and genetic connections (comets and TNOs) forming a virtually continuous population. Taken together, the characteristics and relationships between the various bodies produce the strongest constraints available for modeling the formation of our solar system, understanding the early delivery of water and organics to Earth and thus the origin of life, and understanding the current hazard to Earth from impacting bodies. To enable these studies relating to the ensemble properties of small bodies, a coherent data archive with associated tools and interfaces must be available, and the SBN accommodates and excels at this function.

The SBN has archived data from every NASA and foreign small bodies mission, and we propose to continue that work with the current and upcoming missions Dawn, Rosetta, New Horizons, Hayabusa 2, OSIRIS-REx, and any future small bodies missions that might be selected. SBN mission archiving spans the entire life of the mission, starting with helping to design the archive plan before selection. After a mission is selected, we work with the mission archive lead and other mission personnel to help with the data product and archive design to maximize usability and discoverability, and plan the delivery schedule and peer review schedule to insure both timely archiving and availability of the data and its reliability and scientific integrity. The process is streamlined to insure efficient use of mission and PDS resources, and continues until all mission data have been integrated into a robust and scientifically useful PDS archive.

SBN is unique within the PDS in that only a tiny subset of our archived target bodies, including over 290,000 catalogued objects, will ever be visited by spacecraft. To fully achieve the benefits from the detailed in-situ observations of mission-targeted bodies, comparisons with the rest of the population, via ground-based and space-based remote sensing data and laboratory measurements, is needed. Comparative data are also essential for providing the scientific basis for future missions, identifying future mission targets, and characterizing target properties for mission planning and execution. We will expand our acquisition of small bodies-related data by coordinating with NASA-funded research programs that now require archiving under the ROSES requirements, acquiring targeted data from world-wide asteroid researchers, and procuring relevant laboratory data.

User Services

SBN provides data to the user community of small bodies researchers in a way that fully supports and facilitates planetary science research. In addition to providing a curated archive of the vast diversity of mission and ground-based small bodies data along with full metadata and documentation, we also provide specialized search capabilities to allow researchers to harvest the data they need to support their research. The PDS4 standard has been designed to support sophisticated user access to the PDS data resources, and SBN will capitalize on this design by developing enhanced PDS4 search structures that fully and specifically serve the diversity of small bodies data.

SBN invests serious effort assisting the users with transition to PDS4 standards. SBN representatives organized and held PDS4 workshops at DPS meetings and at local and international institutions. We plan to continue this activity and extend it to a broader range of meetings (e.g., ACM, COSPAR). SBN has created and constantly updates a Wiki page that provides tutorials and answers PDS4 user questions. We are also developing software to allow users to prepare their data according to PDS4 standards (OLAF) and to read and manipulate the PDS4 data (e.g., IDL package READPDS4).

In addition to enhancing PDS4 search services by optimizing the metadata to small bodies data, we have also developed specific user search tools to facilitate user access to our holdings. Our new partner at Bowie State University was brought onboard to expand these capabilities using his expertise in accessing large collections of very diverse types of data. Currently, our Small Bodies Image Browser (SBIB) provides map-based searching and retrieval of data from large mission data sets, which can be downloaded in a variety of formats for available analysis software. Currently, SBIB supports Dawn Vesta and NEAR Eros data, and will be expanded to Dawn Ceres, Hayabusa Itokawa, Rosetta 67P/Churyumov-Gerasimenko, Galileo Gaspra, and others. A complementary tool designed to support research using the data, is the Small Bodies Mapping Tool (SBMT), which projects various data onto shape models. The SBMT will be delivered by APL and installed at SBN during the first year of this proposal. Our Small Bodies Data Ferret is a powerful tool for discoverability in the vast diversity of ground-based and mission-derived data sets. It is a web-based database tool that we developed to integrate data across the entire SBN asteroid, comet, and small planetary satellite data holdings through target-based searching. The Data Ferret allows a user to extract and display data relating to any specific target or list of targets covering over 290,000 objects represented in the SBN data holdings. During the proposal period, the Data Ferret will be expanded to support query based searching, and will be updated to the PDS4 standard as the archive holdings are migrated to that standard.

NASA has placed significant emphasis on the search for Near Earth Objects (NEOs) for threat mitigation as well as the possibility of resource harvesting. The resulting NEO surveys have collected vast amounts of data of great value to the planetary science community and to the public, but these observations are in danger of being lost, as the original surveys are not funded to provide a long-term archive of their data. Archiving of the Near Earth Asteroid Tracking (NEAT) survey is partially completed, and we are working with the Catalina Sky Survey, LONEOS, and Spacewatch surveys to obtain their data as well. To provide easier access to these large databases, we are developing the NEO Survey Search Tool, which calculates orbits of any catalogued solar system object and uses the results to identify all archived survey images that might potentially contain the object.

To ensure the full scientific value of our archives we have developed a robust system of external peer review, subjecting all submitted data to scientific review by top experts in relevant planetary science research fields, as well as thorough internal review for completeness, correctness, and full PDS standards compliance.

SBN directly engages the small bodies research community. Regular reports are given to NASA's Small Bodies Assessment Group for feedback in addition to announcements for input through the SBAG listserv and the Planetary Exploration Newsletter. We have an Advisory Council of prominent researchers that meets annually and is polled by email during the year.

Management

The headquarters of SBN at the University of Maryland (SBN/UMd), led by PI Michael F. A'Hearn, is the primary institution for the management of SBN and is the primary point of contact. SBN/UMd provides overall management of the node and has as its science and archiving focus active small bodies, particularly comets. The SBN division at the Planetary Science Institute in Tucson, AZ, led by institutional PI Eric Palmer, has as its science and archiving focus asteroids, small planetary satellites, and interplanetary dust. The two parts of SBN work cooperatively, coordinating by means of bi-weekly telecons and twice annual face-to-face meetings. To ensure relevance and value to the community, both branches of SBN are staffed with scientific researchers who are separately funded to perform research in the scientific areas related to the archive.

The two institutional branches of SBN, although they are in close communication and cooperation, are geographically and institutionally distinct, and maintain separate IT infrastructure and software development environments. They take advantage of this by mirroring their holdings (the UMd mirror at PSI is installed on-site and the PSI mirror at UMd is fully operational). Independent and collaborative software development projects are undertaken at and between the institutions.

An important part of SBN's function is close communication and cooperation with the other PDS nodes and the technical support nodes. In addition to interaction at Management Council meetings, we support one another's peer reviews, use one another's software tools, and often cooperate in the archive work of individual missions, as with the LADEE mission, where the Atmospheres node is the lead but SBN is leading the archiving for the LDEX instrument, or with Dawn where SBN is the lead but the Radio Science (RS) advisor provides critical support for archiving the RS data. We are currently funded under PDART in a joint project with the Rings Node to make planetary data in NASA's astrophysics archives directly accessible in PDS. We also work with NAIF on all our missions to insure that the supporting geometry and other SPICE information are fully integrated into each mission's archive. SBN's goals are: to provide a fully-supported and curated archive of all available small bodies data, both from missions and from ground-based observations; to provide exceptional user support to maximize utilization of these data for planetary science research; and to pursue research into innovative methods of metadata and formatting standards to improve archive accessibility and the cost-effective use of resources.

Planetary Data System Engineering Node Performance Review

Executive Summary

Overview

The Planetary Data System (PDS) is organized as a distributed federation of science discipline nodes responsible for the archiving and distribution of planetary science data. As a federation, each discipline node is responsible for the capture and management of archival data along with the corresponding software infrastructure and services to support their distinct scientific discipline. *The Engineering Node, located at the Jet Propulsion Laboratory (JPL), provides software and data engineering support to the entire federation of nodes. The Engineering Node plays a critical role in the overall architecture, implementation, and operations of the Planetary Data System, having led the recent development of PDS Version 4 (PDS4), the largest upgrade in the history of the PDS, to an online, distributed, model-driven, service-oriented architecture.* Over the past few years, the structured approach put in place by the Engineering Node for moving the PDS federation towards an international system has been paramount to the successes of operationally adopting PDS4 for upcoming planetary missions. The maturity of PDS4 enabled it be transitioned from a development system to an operationally deployed capability successfully supporting the Lunar Atmosphere and Dust Environment Explorer (LADEE) and Mars Atmosphere and Volatile Evolution (MAVEN) missions. The innovation in PDS4 and the rigorous engineering approach led the PDS4 Operational Review Board in September 2013 to state: “We would like to commend the PDS team on a truly excellent piece of system and software engineering, and recognize that you have figured out how to successfully navigate and manage a potentially very difficult distributed and diverse community.” The community not only includes PDS Discipline Nodes, NASA planetary missions, and users, but also the international community through the International Planetary Data Alliance (IPDA) that has adopted PDS4 and is preparing its use on their upcoming missions. The Engineering Node has successfully worked with the planetary science community to develop and deploy an innovative data science infrastructure that will serve as the data ecosystem for planetary science for years to come.

The PDS Management Council, which includes members from each science and support node within the PDS and provides oversight for the federation, established a mission statement in its 2007-2016 roadmap: “Facilitate achievement of NASA’s planetary science goals by efficiently collecting, archiving, and making accessible digital data and documentation produced by or relevant to NASA’s planetary missions, research programs, and data analysis programs.” To accomplish this, the PDS will: 1) gather and preserve the data obtained from exploration of the Solar System by the U.S.; 2) facilitate new and exciting discoveries by providing access to and ensuring usability of those data to the worldwide community; and 3) inspire the public through availability and distribution of the body of knowledge reflected in the PDS data collection. The development of PDS4 is a direct response to the roadmap efforts established by the PDS Management Council.

The development and release of PDS4 has formed a basis for achieving these goals and to build future capabilities for the PDS and planetary science community. The architectural decisions will allow PDS to scale into the Big Data Era, for both missions and users, beyond the current one petabyte archive, for years to come. This is critical to achieving the vision as stated by the NRC in the Planetary Science Decadal Survey, 2013-2022: “*to support the ongoing effort to evolve the Planetary Data System from an archiving facility to an effective online resource for NASA and international communities*” [1]. The transformation from archiving data towards an international platform to enable planetary data discovery and research that can pull the data across international archives together forms a basis for continued evolution of PDS over the next five years. This includes not only improving access to data in PDS archives, but also enabling the PDS to become a knowledge-base for planetary science. In addition, the extensibility provided by PDS4 will allow PDS to put new science data services in place. The architecture decision, especially the formation of a rich planetary science information model, enables the PDS to move towards forming such a system and integrating data and resources together to support the international community and to support the future mission and research needs. The Engineering Node will play a critical role in achieving this by building out the models, software tools, and infrastructure services, and by working with the data archives within the PDS and IPDA communities, along with other value added data and information, to construct this knowledge-base.

Management of the Engineering Node is led by Daniel Crichton, Principal Computer Scientist and Program Manager at NASA/JPL, who most recently has worked with the National Research Council and the NASA Office of Chief Technologist on developing reports [2] and roadmaps for big data and data intensive systems [3]. He also serves as JPL's leader for the Center for Data Science and Technology, a joint initiative with Caltech, to further the technology and development of data science capabilities to support the capture, management, and analysis of scientific data. Additional leaders at the Engineering Node (J. Steven Hughes, Sean Hardman, and Emily Law) are all actively involved in developing data intensive systems across multiple disciplines and pulling in best practices to develop the PDS.

Future/Long-term Roadmap Actions

In addition to the system drivers identified above that affect PDS for the foreseeable future (next five years), the Engineering Node anticipates, given the increasing volume of data, the distribution of archives, and the varying user needs, that two principle themes will be required. The first, is to ensure that PDS can scale to capture, manage, and archive data from future PDS data providers. The second is to support the varying needs of planetary science users and to ensure they can access data internationally. This is in response to the NRC's statement in the Planetary Science Decadal Survey, 2013–2022: "To support the ongoing effort to evolve the Planetary Data System from an archiving facility to an effective online resource for NASA and international communities." [1].

As identified in Section 4.6, PDS has increased from approximately 100 terabytes to 1 petabyte. Planned missions with high-resolution instruments and improving communications will add to this significantly, and it is critical that the infrastructure and technology investments are kept so PDS can evolve its implementation. This increase in scale, coupled with increasing data producers and international collaborations, requires that the Engineering Node ensure that the PDS and international community has the tools and services to efficiently work with missions to capture the data. From a technical perspective, this will require increasing automation in data deliveries from missions to the PDS (Discipline Nodes, Engineering Node, and NSSDCA), PDS-wide tracking of data through an integrated system, cost effective storage strategies including managing primary and secondary archives copies of data, and distribution and data movement mechanisms to efficiently move data across the PDS as well as to users and the NSSDCA. The Engineering Node anticipates that increasing instrument capabilities coupled with newer communication mechanisms will continue the trend with increasing data volumes.

Longer-term, *major opportunities exist in extending user services and support*. The development of PDS4 and the partnership with IPDA provides an excellent foundation for shifting from individual archives *towards an international platform for planetary science research*. While much of the focus on PDS4 has been working to prepare missions and nodes for adopting and delivering PDS4 data, the increasing availability of data over the next five years in the PDS4 format will provide an excellent opportunity for leveraging the standards and infrastructure put in place to improve the transition towards an integrated international platform. As a service-oriented architecture, the PDS4 platform (services, tools, interfaces, etc.) can be extended to support this increased integration. This includes continued deployment of services at the Engineering Node, Discipline Nodes, and IPDA member archives, to support international search, transformation and operations of and on the data, and APIs to allow users to plug in and access PDS data directly from their tools and applications. The registry and search tools in PDS4 provide an excellent foundation for building the international search capabilities. *Furthermore, the use of open source, modern search technology (e.g., Apache Solr), provides excellent opportunities to tailor both generalized and discipline-specific searches to support a multi-level search strategy (e.g., looking across the PDS, and then digging into specific data sets at nodes).*

In addition to improving access, the service-oriented architecture will also allow for *integration of advanced computing services to operate directly on the data*. These types of services can apply methods on the data to improve data discovery (e.g., through data mining techniques), transformation, and analysis. Given the integrated knowledge-base of an international platform, providing integrated software services that can transform and compute on the data represents a tremendous opportunity for supporting new use cases from the greater science community while operating on the data within repositories. This increasing shift from stewardship towards greater computation on the data is something that will need to be closely coordinated with the Management Council, NASA Headquarters, and the Planetary Science Community to ensure that it fits within the Level 1 requirements of PDS and can support the evolving needs and expectations of the user community. The Request for Information (RFI) from PDS Management on the street regarding user needs should help to identify future capabilities and requirements. Tables 5 and 6 show planned capabilities to shift the focus for the specific data management and archiving functions, as well as the information architecture supported by PDS4 and the Engineering Node, towards greater user support as the platform is realized. Improved data science-related capabilities and new information technologies can be added to increase the effectiveness of PDS over time with the architecture and infrastructure in place.

PLANETARY DATA SYSTEM CHARTER

The Planetary Data System (PDS) assists NASA in achieving its planetary science goals by efficiently collecting, archiving, and making accessible digital data produced by or relevant to NASA's planetary missions, research programs, and data analysis programs.

The Planetary Science Division relies on the Planetary Data System to implement those Goals through its Headquarters Program Executive, Headquarters Program Scientist, and its Project Manager at Goddard Space Flight Center.

The PIs of all of the Discipline Nodes, along with the leaders of the Technical Support Nodes, the Project Manager, and Deputy Project Manager form the PDS Management Council (MC). The MC serves as the Technical Policy Board of the PDS, and provides findings for NASA with respect to planetary science data management, ensures coordination among the Nodes, guarantees responsiveness to customer needs, and monitors the appropriate uses of evolving information technologies that may make PDS tasks both more efficient and more cost effective.

As needed, the Council may convene subgroups, including Technology Subgroups, to pursue specific actions and report results to the Council.

[61] Charter https://pds.nasa.gov/about/pds_charter_12102015.pdf

2015 Request for Information Responses

The RFI was issued in November 2015. And responses were received by January 5, 2016. The table below summarizes the RFI responses and how the RST incorporated their input, but their full text can be found at the URL linked above. The RST randomly assigned a 3-digit identifier number to each response at the URL above, and those identifiers are used in the table below.

RFI Response Identifiers	Concept Summary	Numbered Reference	Reference Description
001, 003	The core archiving mission of PDS is essential and should be maintained	**Section 2.2**	**PDS Requirements**
001, 003, 009, 017, 019, 021	The PDS should archive more types of data	**Finding XIV** **Finding XV**	**Astromaterials Data I** **Astromaterials Data !I**
001, 006, 022	PDS should consider archiving Software to produce data products	**Finding X**	**Archiving Software**
017	PDS should require archiving of Ancillary data like optical distortion and radiometric calibration data	**Not addressed explicitly**	PDS cannot “require” what people archive, but these kinds of data sets are very welcome.
015, 016, 023	PDS should consider archiving curation sample data and cross-section data	**Finding XIV**	**Astromaterials Data I**
004, 007, 008, 010, 014	PDS should provide a generic, modern access ability to all PDS data	**Finding VII**	**Access to Data**
012, 018	PDS should provide a generic, modern search capability to all PDS data	**Finding III**	**Data Discoverability**
011	If PDS archives higher-order data products, then more sophisticated services should be provided for access to that data (e.g. mapservers):	**Finding XIII**	**Higher-order Data Products**
014, 021	Provide more data formats or more translators for users to convert “PDS” data into formats more usable by them	**Finding IX**	**PDS File Formats and Translation Software**
001, 017, 020	PDS should provide software to allow the community to create submission-ready data for the PDS	**Finding XII**	**Potential Impact of ROSES Archiving Requirements**
014		**Finding XII**	**Potential Impact of ROSES Archiving Requirements**

RFI Response Identifiers	Concept Summary	Numbered Reference	Reference Description
	PDS should provide a validator for data that is retrieved from the PDS	**Section 5**	**Conclusion / Summary**
013	PDS should have a way to easily visualize observation geometries and then download the relevant SPICE data	**Section 3.2**	**Data Discoverability**
		Finding VIII	**Documentation and Training**
021, 023, 024	PDS should enable users to share/exchange workflows and synthesize multiple data sets	**Section 2.1**	**What is PDS and What Does It Do?**
017, 021	PDS should provide more training	**Finding VIII**	**Documentation and Training**
007, 017	PDS should use international standards and a common set of terminology	**Finding VI**	**Modernizing Metadata**
021, 024	PDS should integrate with other national data repositories	**Finding IV**	**Integration with Other Archives**
005	PDS should consult with the MAPSIT group on cartography-related issues	**Not addressed explicitly**	Relevant Node(s) are expected to interact with their communities and interaction with MAPSIT is an example of that, since this interaction is already a part of PDS procedure the RST did not address this specifically.
002	Advertisement for services	**Not addressed**	The RST not find applicable material in this response relative to its charter.

DATA MANAGEMENT, PRESERVATION, AND THE FUTURE OF PDS

Reta Beebe - New Mexico State University, Las Cruces NM Telephone: (575) 646-1938
Email: rbeebe@nmsu.edu

Co-Authors
Acton, Charles - Jet Propulsion Laboratory, Pasadena CA
Arvidson, Raymond - Washington University, St Louis MO
Bell, Jim -Cornell University, Ithaca NY
Boice, Dan - Southwest Research Institute, San Antonio TX
Bolton, Scott - Southwest Research Institute, San Antonio TX
Bougher, Steven -University of Michigan, Ann Arbor MI
Boynton, William -University of Arizona, Tucson AZ
Britt, Daniel -University of Central Florida, Orlando FL
Buie, Marc - Southwest Research Institute, Boulder CO
Burns, Joseph - Cornell University, Ithaca NY
Capria, Maria Teresa – IASF-INAF-Roma/Past chair of IPDA, Rome IT
Coradini, Angioletta - IFSI-Roma/PI of Juno/JIRAM, Rosetta/VIRTIS & Dawn/VIR Rome IT
Crichton, Daniel - Jet Propulsion Laboratory, Pasadena CA
Ford, Peter - Massachusetts Institute of Technology, Cambridge MA
French, Richard - Wellesley College, Wellesley MA
Gaddis, Lisa - U.S. Geological Survey, Flagstaff AZ
Gierasch, Peter - Cornell University, Ithaca NY
Gladstone, Randy - Southwest Research Institute, San Antonio TX
Gordon, Mitch - SETI Institute, Mountain View CA
Greeley, Ronald - Arizona State University, Tempe AZ
Hansen, Kenneth - University of Michigan, Ann Arbor MI
Jakosky, Bruce - University of Colorado, Boulder CO
Kasaba, Yasumara - Tohoku University/Current Chair of IPDA, Sendai City, JP
Khurana, Krishan - University of California Los Angeles, Los Angeles CA
Kurth, William - University of Iowa, Iowa City IA
Law, Emily - Jet Propulsion Laboratory, Pasadena CA
Lorenz, Ralph - JHU Applied Physics Lab, Baltimore MD
Nixon, Conor - Goddard/Univ. of Maryland, Greenbelt. College Park MD
Paranicus, Chris - JHU Applied Physics Lab, Baltimore MD
Pryor, Wayne - Central Arizona College. Coolidge AZ
Roatsch, Thomas - DLR Institute of Planetary Research, Berlin DE
Russell, Chris - University of California Los Angeles, Los Angeles CA
Schwehm, Gerhard - European Space Agency, Villa Franca ES

[62] Submitted as a White Paper as part of the Community Input for the 2013 – 2022 Planetary Decadal Survey (http://www8.nationalacademies.org/ssbsurvey/publicview.aspx)

Simpson, Richard - Stanford University, Palo Alto CA
Sykes, Mark - Planetary Science Institute, Tucson AZ
Tholen, Dave - University of Hawaii Oahu, HI
Walker, Raymond - University of California Los Angeles, Los Angeles CA
Withers, Paul - Boston University, Boston MA
Zender, Joseph - European Space Agency, Noordwijk NL

PREFACE

Efficient, effective archiving and distribution of data is an integral part of planetary science research. We strongly encourage the decadal committee to support the concentrated effort currently underway to evolve the Planetary Data System (PDS) from an archiving facility to an effective on-line resource for the NASA and International communities (the PDS2010 project).

We urge the committee to incorporate the following three issues in the final report:

- Identify as a high priority, the need for broad emphasis from the NASA Planetary Science Division to assure that its policies and procedures guarantee adequate, consistent support for data analysis within the missions and the community and to enable effective archiving.
- Strongly recommend that future NASA Planetary Science Division NRAs and AOs include specific requirements that in addition to raw data, missions and instruments provide data in physical units. Archive planning should be an integral part of the proposal planning, and funding should be identified in the award to ensure teams have adequate resources to meet this additional obligation.
- Strongly recommend that the NASA Planetary Science Division support the upgrade of PDS including leveraging modern data base and Web 2.0 technologies in order to ensure improved data standards and efficient, effective storage, search, retrieval and distribution of scientifically useful planetary data in the coming decades.

Background and Justification

1. Introduction

Planetary exploration by spacecraft represents significant national investments that cannot be easily repeated. Return visits by more capable spacecraft are rare and depend on results obtained from precursors. Events observed by Earth-based or *in situ* instruments often unfold slowly and are not repeatable. Scientists need access to original data to verify reported results, to test new insights and theories, to carry out time-dependent studies, and to assess limitations in our knowledge so that future observations can be planned.

If an archive is comprehensive, readily accessible, and usable, it can meet the needs listed above — it can serve as a virtual reflight of missions and observing campaigns preserved in its contents, at a cost which is minuscule compared with acquiring the original data. However, creating and maintaining a high-quality archive requires commitment from the funding agency, the data providers, and the users—a point which was recognized by the National Academy of Science in its mid-decadal "report card" (ref.: Grading NASA's Solar System Exploration Program: A Mid term report – Co-chairs W. Huntress and N. Noonan (2008) ISBN:0-309-11493-4).

In the remainder of this paper we discuss the present state of data management and archiving within the Planetary Science Division and our recommendations for improvement within the PDS 2010 framework.

2. Background

Although NASA had been including language in contracts for several years that required data from planetary missions be submitted to the National Space Science Data Center (NSSDC), it was clear by the 1980s that a more methodical process with better user access was needed. The NSSDC collection was important as a deep archive, but, because of the lack of a process that provided direct interaction of mission teams and qualified scientists for assessing development of the products, its contents were highly variable in terms of both quality and content. Viking and Voyager stimulated interest both in 'data mining' (searching acquired but previously unexamined data) and reanalysis (seeking new discoveries from previously studied data) and the demand for direct access to mission products increased.

After a study and a prototype phase, the former Solar System Exploration Division (SSED) at NASA established the PDS in 1989. The PDS was a distributed system, with a central node (incorporating both management and engineering functions), supporting nodes (imaging and the Navigation and Ancillary Information Facility (NAIF)) and discipline nodes (DNs) responsible for science data at home institutions that qualified as 'centers of excellence' in atmospheres, geosciences, particles and fields, rings and small bodies. Creation and structure of PDS were based on recommendations from the National Academy of Science (NAS) Committee on Data Management and Computing (CODMAC) (1982, 1986 and 1988) that archives should be housed with science expertise (and, within the context of dealing with multiple short-lived missions this recommendation has proven to be a workable solution for integrating science discipline expertise into the archived products). It was established that PDS would explicitly serve the SSED-funded community and NSSDC would receive copies of PDS data sets for permanent archiving and distribution to non-NASA researchers, international scientists, educators, and the general public. Early data transfers were by magnetic tape and later by CD and DVD physical media, which led to the use of ISO9660 compatible structure and naming conventions that are still in use.

PDS formation was roughly coincident with the birth of the world wide web. In the ensuing two decades, improved technologies produced ever increasing data complexity and data volume while network communication transformed both how PDS did its business and how it interacted with both data providers and data users. The 'distributed' system that was designed for dataset exchange via tape or CDs was integrated so that queries for data could be submitted not just from home institutions but from personal computers from homes, hotel rooms and foreign shores. Instrument teams began delivering terabytes of raw and partially processed data; calibration files were continually being revised, leading to new versions of higher-level products. And users began asking for not only more support, but more sophisticated support—*could PDS provide all* *of the atmospheric temperature-pressure profiles over the PHOENIX landing site, could images from the Shoemaker-Levy 9 Jupiter encounter be recalibrated, and did near-infrared spectra exist (from any source) of asteroid 4370 Dickens?*

The PDS evolution initiated by the web revolution has been uneven and strongly constrained by a limited budget. With the significant increases in data volumes (Figure 1), the challenge of capturing and protecting the bits themselves is daunting. On the other hand, the process for assuring the long-term integrity of the data has become more generally tractable – PDS data are now monitored and distributed with checksums, replicated at mirror sites and each node has a backup and recovery plan.

While modern web services and the more complex needs of the science user communities have produced substantial increases in expectations with regard to high granularity access and highly processed data, leverage to assure uniform delivery of data from instrument teams has been severely limited. For example, some instrument teams want to deliver data in a range of processing states; others are content to make their raw binary files public and let others generate higher-level products. PDS representatives work with instrument teams, often with widely differing approaches to data archiving, to ensure as much uniformity as possible in presentation at each processing level and to produce documentation that is understandable to users familiar with the field. The situation has been hampered by NASA's earlier inattention to enforcing delivery of calibrated products and to assuring that adequate funding was reserved by the missions to allow the teams to produce standard products, especially at the highest processing levels, which are in greatest demand.

Another issue that the PDS confronts is calibration. For some instruments, the calibration of data is an ongoing process that can take years or longer. The reasons for this vary and include such diverse issues as the accumulation of sufficient data to draw proper conclusions to working with flight spares to revisit calibration issues. Data analysis is a process. This is why raw data is often not useful and calibrated data can often be eclipsed. The resolution of this problem is not to offer nothing, as has been the case in the past, but provide cautions and to offer intermediate products which are the best products that can be reasonably provided at a given time, and to provide for a final set of calibrated products once the calibration process has been stabilized or at end of mission.

In 2005 NASA reorganized the PDS, moving the management to Goddard Space Flight Center and reorganizing the JPL-based engineering node. At the same time NASA became more vigilant in requiring that missions plan and budget for data analysis and archiving.

3. The Diversity of Planetary Science

Planetary data are acquired with flyby and orbiting spacecraft making both remote and in situ measurements, surface stations, rovers and sample return missions. Mission lifetimes range from months to decades. Archiving this diverse reservoir as well as supporting ground-based observations, laboratory data and spacecraft radio tracking and engineering information is challenging. The need to apply standards that assure long-term preservation and data integrity imposes additional constraints on PDS policies and procedures. PDS differs from a facility such as the Space Telescope Science Institute (STScI), which deals with an accumulating archive from a few very specific instruments. HST has developed a data pipeline and can provide an on-the-fly calibration service based on the latest and best calibrations requested by users. However, in the course of its nearly two decades of operation, HST has obtained data from 18 remote sensing instruments. In contrast, PDS works with a much more diverse set of instruments and teams where virtually all planetary exploration 'observatories' are born and die on time scales that are short when compared to HST's operating lifetime.

During the first half of 2009, PDS ingested data from 82 instruments on 17 spacecraft ranging from three-dimensional *in situ* magnetometer data to gigabyte image strips from push-broom cameras. By 2015 the PDS is projected to house data from 511 instruments from 70 spacecraft. This will result in an estimated data volume of 245 terabytes from the individual data sets included in Figure 1.

Each planetary mission defines its observations, collects its data, and deposits its results in the archive within a few years. Funding disappears before calibration on many instruments is fully mature. In addition, lack of oversight and mission funding to produce higher-level products from the wide range of instrumentation, and divergent community practices among disciplines have led to PDS data sets that are not easily compared with each other and (in some cases) poorly understood except by those who were involved in the data acquisition. The reality is that PDS must support many different data pipelines each optimized for its mission and instrument.

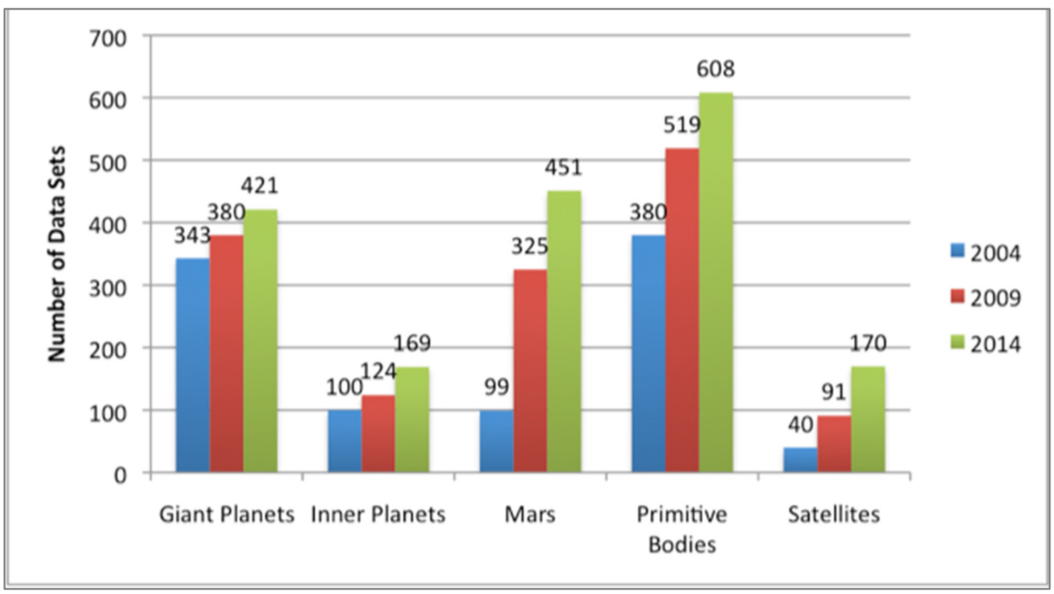


Figure K.1. The Estimated Number of Accumulated Data Sets Per Decadal Category. Note: Lunar missions are included in the satellites category. Possible contributions from the following missions have not been included due to lack of information or because they are scheduled beyond 2014 -- **Satellites:** Grunt (Russia) Phobos, Yinghuo-1 (China) Phobos, Chang'e 1 (China) Lunar, Chang'e 2 (China) Lunar, Kaguya (Selene) Lunar, SELENE 2 (Japan) Lunar, Chandrayaan 1 (M3, MINI-RF) Lunar, Chandrayaan 2 (India) Lunar, Luna-Glob (Russia) Lunar, Lunar mini-Landers Lunar, MoonNEXT (ESA) Lunar, MoonLITE (UK) Lunar, Smart-1 (ESA) Lunar, LAPLACE (Ganymede)(ESA) – **Mars:** ExoMars (ESA) Orbiter, ExoMars (ESA) Lander, MarsNEXT (ESA), Mars Sample Return – **Inner Planets:** Bepi Colombo (ESA), Venus Express, Venus Climate Orbiter (JAXA) – **Giant Planets:** Outer Planet Flagship (launch 2016) – M**iscellaneous:** Discovery AO-2008/9, Discovery AO-2010, New Frontiers 3.

To access a summary of timelines for missions in operation and under development see http://atmos.nmsu.edu/pub/download/NASA_Mission_Summary_special_022409.xls

4. International Implications

The Decadal Survey represents an opportunity to raise awareness of the rapidly changing data management and archiving requirements for the next decade(s) and to recognize the growing trend for international cooperation on missions and in data sharing. Along with the growing internationalization of space comes an urgent need both to ensure the preservation of and to provide access to an ever-increasing volume of usable planetary data worldwide. When ESA began plans to establish their Planetary Science Archive in the mid-1980s, influenced by the need for rapid progress and the fact that there was already an experienced ESA cadre of PDS users, they adopted PDS data standards. Subsequently, when India's ISRO selected NASA and ESA instruments for Chandrayaan-1, they adopted PDS standards for their archive. As a result, ESA and NASA worked together to establish the International Planetary Data Alliance (see http://planetarydata.org) in 2006 as a mechanism to develop international standards for planetary science data archiving and encourage international interoperability. Japan's JAXA has accepted PDS standards and is working through the International Planetary Data Alliance to adopt the interoperability protocol developed by ESA/NASA for access to Venus Express data for their Venus mission, Planet C. At the same time, China's CNSA is developing Chang'e-1, and individuals are working to establish archives that will be PDS compatible. Thus, the PDS standards have become the de facto international standards and, although these agencies are receptive to PDS leadership, it is the responsibility of PDS to strive to produce well-defined standards to sustain the efforts to make archives from international missions available in compliant formats to all users.

Improvement of the PDS and international access can yield significant benefits to the planetary program by ensuring that best use is made of data collected in past and ongoing explorations. A dynamic model for data archiving and management within NASA is an essential component in planning our role in the future of space exploration. Only by supporting continued improvement of the PDS can NASA capitalize on data collected in past and ongoing explorations.

5. Expectations of the PDS2010 Project

In coordination with international archiving agencies through the IPDA, PDS will:

- Develop revised, rigorous but simple archiving standards that are consistent, easy to learn, and easy to use;
- Accept a limited number of archive data formats, which will simplify development of data management, conversion, and manipulation;
- Provide adaptable tools to both mission and ground-based data suppliers for designing archives, preparing and validating data, and optimizing delivery to the PDS;
- Develop a standard interface with the Solar System Exploration R&A and DAP programs that assures that participating scientists who have proposed to deliver data can do so in the most efficient and effective manner;

- Leverage modern web and computing technologies to support the operations as a fully online, distributed, international data system;
- Provide better access allowing users to identify, transform and obtain selected data quickly from anywhere in the system;
- Work with the instrument teams to develop tutorials for data that are intrinsically difficult to use;
- Flag intermediate data that are to be used with caution while instrument teams do ongoing calibration work;
- House and provide access to models that have been developed by the science community and in common use;
- Provide a highly reliable, scalable computing infrastructure that protects data integrity, links Data Nodes into an integrated data system, and provides the best service to both data providers and users for at least the next decade.

6. Responsibilities

Recent attention at NASA Headquarters to the need for more clearly specified proposal and mission requirements and reorganization of the management structure of PDS has set into motion processes, which are leading to considerable progress in achieving the goals above.

There are many stakeholders and each has responsibilities that must be clearly identified and supported by NASA in order to ensure successful data archiving and access. Those stakeholders include NASA Headquarters, the Planetary Data System, Principal Investigators of PI-led missions or Instrument PIs on Flagship or Other Missions, Ground-based Suppliers of Telescopic and Laboratory Data and the Data End User Communities. We have attempted to enumerate these responsibilities in a draft charter for archiving (see http://atmos.nmsu.edu/pub/download/Planetary_Science_Draft_Charter.pdf). However, none of these requirements can be reasonably addressed unless NASA Headquarters assigns sufficient priority to the requirements for archiving and funding to allow teams to analyze the data sufficiently to complete calibration and documentation so that all the stakeholders can meet the requirements our preliminary archiving charter defines.

7. Conclusions

Even though PDS received high marks in "Grading NASA's Solar System Exploration Program –A Mid Term Report", Co-chaired by Norine Noonan and Wesley Huntress, it should be noted that this was a progress grade that was based on current Headquarters approaches and PDS progress since reorganization in 2005. If this momentum is to be sustained at a level that will allow the PDS to transform into the online research support facility that will serve the science community to make optimal use of mission data, the Planetary Science Division must continue to stress the importance of end-to-end management of data acquisition, adequate funding for data analysis within the missions and in data analysis programs, and completion and maintenance of PDS2010 to ensure PDS will meet the solar system exploration challenges of the next decade and continue providing improved user services.

National Aeronautics and Space Administration

Goddard Space Flight Center
8800 Greenbelt Road
Greenbelt, MD 20771
www.nasa.gov/goddard

www.nasa.gov

NP-2017-6-024-GSFC