The Planetary Data System:
A Vital Component in NASA's Science Exploration Program

Authors

Louise M. Prockter, *JHU/APL* (louise.prockter@jhuapl.edu 240-639 5042)
Matthew S. Tiscareno, *SETI institute*
Edwin J. Grayzeck, *NASA GSFC*
Charles H. Acton, *JPL*
Raymond E. Arvidson, *Washington University*
James M. Bauer, *University of Maryland*
Reta Beebe, *New Mexico State University*
Sebastien Besse, *ESAC*
Nancy Chanover, *New Mexico State University*
Daniel J. Crichton, *JPL*
Lisa R. Gaddis, *USGS*
Mitchell K. Gordon, *SETI Institute*
Trent M. Hare, *USGS*
David M. Hollibaugh Baker, *NASA GSFC*
John S. Hughes, *JPL*
Emily S. Law, *JPL*
Myche McAuley, *JPL*
Tim McClanahan*, NASA/GSFC*
Jordan H. Padams, *JPL*
Anne C. Raugh, *University of Maryland*
Mark R. Showalter, *SETI Institute*
Thomas Stein, *Washington University*
Ray J. Walker, *UCLA*

Cosigners
See list of cosigners at https://seti.org/PDS-WP20_supporters

1. Executive Summary

The Planetary Data System (PDS) plays a vital role in the planetary science community, enabling an extraordinary scientific return from NASA's exploration endeavors. Led by scientists who serve as a bridge between the data and the research community, the PDS has become the primary platform for planetary researchers to obtain current and historical data, while retaining its original mission of preserving the integrity and usability of data for future generations. In order to facilitate effective stewardship and use of its diverse holdings, the PDS has developed and is implementing the PDS4 Information Model (the PDS4 IM). **In the coming decade, continued strong support from NASA HQ is essential**:

(1) to ensure the complete implementation of the PDS4 Information Model and the development of effective tools that realize its potential for data discovery and data access;

(2) for the PDS to obtain the highest-quality data and documentation from science teams, including detailed user guides and access to internal team tools developed beyond the scope of originally planned data products; and

(3) to accommodate the ever-increasing level of PDS-supported user services, and provide archiving resources for a large number of derived data providers.

2. The Vital Role of the PDS

The PDS's expert curation of archived data serves several purposes: 1) the long-term preservation of data for future generations, 2) ensuring discoverability, accessibility, and usability of the data, 3) supporting data providers in the creation of high-quality archival data, 4) supporting data users in obtaining and understanding the data they need, and 5) providing leadership within the planetary data archiving community, including promoting the interoperability of data archives across NASA and among international planetary archives.

The PDS works with missions to ensure delivery of validated and high-quality archives. Data submitted to the PDS undergo rigorous peer review before they are released as archive products. Peer review ensures the long-term quality and usability by validating data structures and scientific content, often by identifying and correcting errors. The PDS also provides value-added services including generation of additional metadata,[1] more accurate geometric mapping, and coordinate systems transformations and translations between common data formats.

3. The PDS today

NASA established the PDS in 1989 as the scientifically guided archive for planetary science, in order to address concerns that the data being returned by scientific missions was in danger of being lost and/or becoming unusable. The original design of the PDS was guided by the recognition that archiving and curation should be carried out by the scientists who care most about the data [Arvidson et al., 1994]. In that spirit, the PDS adopted a federated structure of nodes organized around science disciplines, which has remained remarkably resilient through the growth in the archive over the past 30 years. This structure has been able to adapt to a remarkable variety of planetary missions, which encompasses diverse targets and environments (solid and gaseous planets, moons, rings, comets, asteroids, KBOs, dust, and charged particles)

---

[1] Metadata is "data about the data", fields and parameters that describe the contents of PDS data products and the observational conditions under which they were obtained.

and diverse platforms (flybys, orbiters, landers, rovers, impactors, probes, etc.). At the same time, the PDS continues to adapt to the latest technological achievements in networking, computational power, databases, storage media, storage density, and accessibility.

3.1 Organization

The PDS is a distributed federation consisting of six Discipline Nodes (DNs) managed by subject matter experts. Each DN operates its own data archive, manages its own data and holdings (Figure 1), and supports a dedicated user community. The DNs are supported by dedicated Data Nodes and Subnodes distributed within the science community.

In addition to the DNs, there are two Support Nodes, both at the Jet Propulsion Laboratory: the Engineering Node, which develops and maintains software infrastructure for archive support, and the Navigation and Ancillary Information Facility (NAIF), which develops and maintains the SPICE geometry system [Acton et al., 2017]. A Radio Science function, currently shared between the Ring-Moon Systems and Planetary Plasma

Figure 1: Current organization of the PDS. Many nodes also house subnodes or data nodes. Discipline nodes are shown in blue; support nodes in tan.

Interactions nodes, supports data providers and users with regard to the specialized needs of radio science. The PDS Project Office is located at NASA Goddard Space Flight Center, and the PDS Chief Scientist is a member of the planetary science research community.
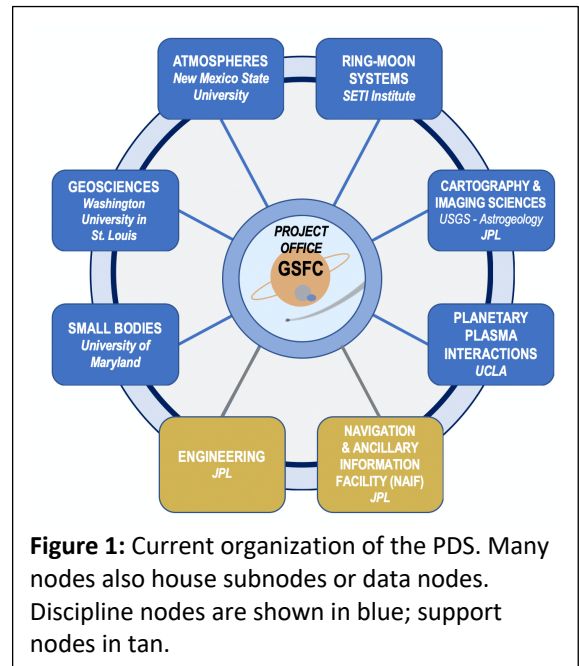
3.2 The PDS4 Information Model

In response to advances in information technology and the ongoing rapid expansion in the volume and complexity of planetary data, the PDS pioneered a scalable information model (IM) suitable for planetary data archives [Crichton et al., 2018; Hughes et al., 2018]. This system, referred to as the PDS4 IM or simply as PDS4, provides rigorous structure and definition of metadata, simplified file structures, and (unlike its predecessor PDS3) is designed to remain functional for decades to come. Through consistent metadata standards that improve data interoperability, the PDS4 archive structure more readily establishes connections between related domestic and international archive holdings and adds new metadata that make data searches more effective. The PDS4 IM and its use of extensible markup language (XML) improves data accessibility and usability by allowing PDS data to have greater operability with the wealth of software and tools available for the development of online, web-based systems. As a result, the PDS4 IM is intended to serve as the foundation of an integrated data services platform that will enable ease of access, search, use, and analysis of worldwide planetary data.

In support of this vision, the PDS is currently working to migrate its current holdings from older data standards into the PDS4 IM standard and to build associated tools necessary to realize the benefits of the PDS4 architecture including cross-platform search. The PDS also undertakes community outreach, conducts training classes in PDS4 requirements, and provides assistance to data providers from spacecraft mission teams and from research and analysis (R&A) proposal teams.

3.3 International leadership

The International Planetary Data Alliance (IPDA) endorsed the PDS4 standard as the format to be used for archiving scientific products for planetary missions. More than a dozen international or national space agencies, including the European Space Agency (ESA) [Besse et al., 2017], the Japan Aerospace Exploration Agency (JAXA), and the Indian Space Research Organization (ISRO), currently use the PDS4 IM to archive data from their own missions, and collaborate with the PDS on the IM's continued development. Because planetary missions in the current era are rarely carried out by just one space agency, international collaboration among archiving authorities ensures that data users will benefit from coordinated archives, enabling them to perform the best possible research.

3.4 Connections to other archives

There are direct connections among many archives within and external to NASA which include the PDS, and serve as critical links within the scientific data ecosystem. Astrophysics archives depend on the PDS's Minor Planet Center (MPC) to identify moving objects within their frames. Members of the exoplanet community use PDS data node search services to identify planetary mission observations suitable for use as exoplanet analog observations. Many connections feed into the PDS as well: DNs host shared planetary data from a variety of non-NASA archives and sources, like ESA's Planetary Science Archive, Japan's JAXA, and data from assets supported by the National Science Foundation. For example, PDS holdings also include cross-divisional data of planetary objects from Astrophysics mission sky surveys like TESS and WISE, and telescopic observatories like SST, HST, and SWIFT.

3.5 Data volume, diversity, complexity, and number of data providers

The PDS archives and makes available observations and experimental data from over 60 years of NASA-based exploration of target types including comets, asteroids, rings, moons, dust, fields, charged and neutral particles, and planets. Instrument types archived within the PDS include imagers, spectrographs, gravity science, altimeters, dust detectors, magnetometers, plasma instruments, and more. Observing platforms include Earth-based telescopes and laboratory workbenches as well as interplanetary rovers and surface stations, not to mention orbital remote sensing and *in situ* payloads. Although some other archives have higher data volume (notably EOSDIS, the archive for NASA Earth Science), the PDS far outstrips other NASA archives in terms of data diversity and complexity (see Table 1).

	Datasets	Volume	Target types	Targets	Instrument types	Instruments
EOSDIS	11,000	17.5 PB	1	1	9	148
PDS	2,000	~2 PB	10	>600	65	700

Table 1: As catalogued in the 2017 PDS Roadmap Study (McNutt et al., 2017), PDS archives are considerably more complex than those within EODIS.

Today, 11 missions are actively delivering data to the PDS in both PDS3 and PDS4 formats. These include Juno, InSight, LRO, MAVEN, MRO, MSL, New Horizons, Mars Odyssey, OSIRIS-REx, Rosetta, and Voyager. Since its inception, the PDS has archived over 1.85 PB of data from over 70 missions, over 4500 data sets from over 700 unique instruments, based on over 100 different investigations using more than 65 different types of instruments, and in over 100 coordinate systems.

Furthermore, the data volume stored by the PDS from planetary missions is predicted to almost double in size (to ~3200 TB) by 2025, just halfway through the next decade. Plans to

incorporate additional holdings from NEO programs including ATLAS and Pan-STARRS are predicted to increase the archives over 3x further to 11,300 TB in 2025. The extreme diversity of data produces additional layers of complication to all aspects of PDS archiving: curation, search, distribution, and long-term preservation.

3.6 Research and Analysis Data

In 2015, NASA began to require Research and Analysis (R&A) proposers to include a Data Management Plan (DMP) in their Research Opportunities in Space and Earth Sciences (ROSES) proposals. Beginning in 2020, DMPs are included in the evaluation of a proposal's merit. Not every DMP involves data being archived in the PDS, but many do. This has required heightened effort from the PDS to work with hundreds of individual proposers and awardees, and has further increased the diversity and complexity of the PDS' archive holdings. Fortunately, the PDS4 IM provides a robust framework for curation and discovery of these data sets.

3.7 Laboratory-Based and Physical Samples

A great deal of progress has been made in archiving laboratory and field-based data

Figure 2: Growth in number of R&A data providers to the PDS since the last Decadal Survey.

sets, including digital data of physical samples, since the last Decadal Survey in 2013. These provide important supporting information for helping to interpret mission-related observations, in addition to enabling new discoveries within the realm of planetary sciences, habitability, and life. Dozens of researcher-supplied data sets are currently being ingested using PDS4 standards, such as the PDS Geosciences Node Spectral Library, which now permanently archives thousands of laboratory spectra of planetary samples traditionally hosted by Brown University's RELAB facility.
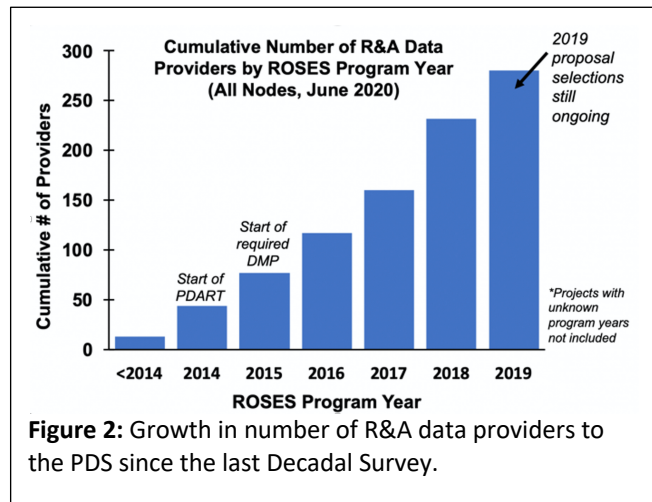
4. Opportunities and Challenges for PDS

4.1 User needs and expectations

Over the past two decades there have been marked changes in what the Planetary community expects of the PDS, in step with broad societal changes in the world's information economy. At the dawn of the 21st century, the focus of the PDS remained on preserving and distributing mission data as whole collections. A researcher might request that a mission's entire dataset be shipped to them by mail, in the form of dozens of CD-ROM disks. The job of determining which of a data set's individual observations were relevant to a researcher's work would be done after the role of the PDS had been completed.

After the information revolution brought about by the rise of Google and related technologies, people now routinely acquire very specific pieces of information with very little time or effort. This new reality has had a profound effect on the PDS, which is now commonly expected to quickly and effortlessly connect researchers with the specific pieces of information that they seek, often for interdisciplinary studies. The PDS embraces this expansion of its responsibilities, but it is important to understand the challenges involved.

4.2 Data discoverability

The PDS3 metadata system, designed for an age when datasets were distributed in their entirety and users requested data by the (known) mission and instrument, could not evolve to keep up with the expanding PDS holdings, the significant increase in the size of individual data products, and the increasing desire of users to mine the archives for data and cherry-pick the results. This was a primary motivation for the creation of the PDS4 IM, which is designed with the new paradigm firmly in mind. Its rich, and extensible, metadata is recorded at all levels of the archive hierarchy, enabling cross-referencing of data at a level unheard of under PDS3. This provides a deep and solid foundation for the PDS to create a new, cross-disciplinary user interface to the data. However, work remains ongoing to migrate existing holdings into the PDS4 IM, to develop new and improved existing metadata, and to implement a full set of tools to exploit the IM's potential.

The keystone of effective search is accurate and consistent metadata across data sets. The planetary research community has a unique focus on both historical data and highly diverse data. Many legacy data sets were not created with cross-discipline search in mind and lack the metadata to support it. The PDS is working to "retrofit" these data sets, but much work remains. Stakeholders sometimes ask why the PDS "cannot be more like Google." The reality is that Google is not so much a search engine as a ranking engine, whose primary function is to answer a query with results that are popular. Such a methodology does not work well for archival data discovery, in which it is necessary to give comprehensive answers to queries with specific, and often highly technical, parameters.

Work is already underway to extend the existing metadata to include the fields needed for integrating the archive, via the Digital Object Identifier (DOI) ecosystem and FAIR data practices, into the greater body of the scientific literature. This integration will make the archive data discoverable through new interfaces, satisfying more users and raising the profile of the data, our data suppliers, and the PDS.

The PDS must ensure that all users have clear and easy access to the best data, regardless of their physical location, to maximize the science return of planetary missions. This includes moving toward a single-point high-level search for scientific products using the PDS standard to maximize the discoverability of science products. Several ongoing activities by the PDS are underway to improve discoverability and accessibility of data within PDS archives, and community workshops [e.g., Gaddis et al., 2012] are held annually to teach users about the latest data, tools, and services.

4.3 Data usability and file formats

Once users have identified the PDS data products that are of interest to them, the focus shifts to the ease of understanding, interpreting, and using the data. There are challenges: data providers and data users commonly prefer to use file formats based on compatibility with currently available software tools; PDS refers to these as contemporary formats. However, because of the PDS' mandate to preserve both integrity and public access to the data into the long-term future, PDS4 standards require archival file formats that are simple and relatively few in number (Hughes et al., 2018). While some contemporary file formats are PDS4-compliant, many are not. In order to bridge this divide, it is essential that data providers and data users have easy access to tools that can conveniently transform between archival file formats and contemporary file formats, and that those tools evolve as new non-archival formats are introduced among the user community. In some cases, such tools are available through the PDS or through third-party developers. In other cases, development of such tools remains needed.

Data usability is further complicated because some legacy data sets have significant shortcomings in documentation or even in the data itself. In some cases, active curation of the data by the scientists at the PDS has included repairing data and/or metadata, or the creation of new or improved user guides. In some instances, the repairs are being made during the migration of the legacy data to the PDS4 IM. Such efforts are ongoing.

4.4 Preserving the legacy of spacecraft missions

The quality of the data in the PDS archives is compromised in many cases by a disconnect between the data analysis conducted by a mission science team and the data set deposited by that team into the PDS. Some data sets currently held by the PDS were forked at an early stage, such that they were never actively used by the mission teams that produced them. The PDS has found that for the data delivered to PDS this often leads to abundant errors and poor usability.

Furthermore, mission science teams often find that certain data products, structures, and representations are useful to them as their work proceeds, but were not envisioned in the pre-mission archiving plan. In some recent cases (notably Cassini), NASA, the PDS, and the mission team worked proactively to capture these products so that current and future generations of researchers may benefit from them.

High-quality documentation is also an issue for some data sets within the PDS. In some cases, PDS scientists have gone to great effort to reconstruct instrument workings and data pipelines, so as to produce needed documentation, but it is much better if the original data providers do this work. Best practice would be for instrument user guides to reflect lessons learned by science teams during the mission, and also to be available in a timely manner so that researchers outside mission teams are able to readily understand and use the data.

4.5 Online processing and analysis

For some data sets, the PDS applies online data processing and analysis mechanisms in service of better usability and exploration. Usability is improved via data transformation and access standardization. As user tools evolve, so must the data and its access evolve to become "analysis ready" and immediately useful to tools deployed in disparate locations. Calibration improvements, geometric mapping, metadata, and coordinate systems transformations are but a few of the basic processes applied to data as new knowledge is discovered. Additionally, use of standardized Application Programming Interfaces (APIs) by DNs allows for more automation and support for interdisciplinary relationships of data types.

Tools to support data exploration help users make organizational sense of the tens of millions of PDS products and filter to more manageable subsets of the data — helping to find the needle in the digital haystack. One approach, Machine Learning (ML) analysis, has the potential to identify features, and produce improved metadata parameterization to extract more and better descriptive information. In some cases, PDS uses several ML classifiers to recognize and locate many planetary and spacecraft features (craters, rings, wheels, etc.) to aid in search and selection mechanisms, thereby allowing users to more quickly focus on data that is pertinent to their investigations.

4.6 Data services

Through the PDS4 IM, the PDS provides a standard model for stewardship of planetary data that has been adopted world-wide. By leveraging technologies including cloud computing, AI, machine learning, data fusion, and visualization, as well as community-developed analysis tools and planetary data services, the PDS should be able to significantly increase discoverability and automation of data

analysis as it virtually unifies data archives world-wide. To realize this vision, the PDS has launched an open Planetary Data Services initiative for the community to increase discoverability from planetary archives, support increased access to data and computation for data analysis, and provide a foundation towards a planetary data ecosystem with other planetary archives through the IPDA.

5. Recommendations

The anticipated challenges that will be faced by the PDS over the next decade cannot be met with level funding, or without support from NASA HQ to ensure policies are in place to lead to good archives. Furthermore, missions should be adequately resourced throughout their lifecycles (including extended missions) to ensure all relevant products and supporting data are delivered to the PDS. We recommend the following:

1. Ensure the complete implementation of the PDS4 IM and the development of effective tools that realize its potential for data discovery and data access.

The full implementation of the PDS4 IM remains in progress. Work must continue to "retrofit" existing data sets to support effective cross-discipline search, including the generation of consistent and high-quality metadata (§4.2). Work must also continue to create effective tools enabling both data discovery and conversion to/from users' preferred file formats (§4.3).

2: Obtain the highest-quality data and documentation from science teams, including detailed user guides produced after the teams have worked with their data, and access to internal team tools developed beyond the scope of originally planned data products.

Archiving should be baked into mission operations from start to finish. The internal data pipelines used by spacecraft mission teams should be well integrated with the pipeline that produces archive products for the PDS (§4.4). PDS needs NASA, and the mission and research communities, to hold mission teams and other data providers accountable to follow best practices in order to ensure the high quality of data ingested into the PDS.

3: Ensure resources for the PDS are sufficient to support the increased demand for user services and archiving resources for a large number of ROSES data providers.

PDS is following the progress of technology, considering costs as well as capabilities, and is refining its Data Services accordingly (§4.5, §4.6). Serving the large influx of data providers from R&A projects funded by ROSES (§3.6) is a challenge, but a welcome one. As this demand upon PDS resources increases, support from NASA should increase apace.

6. References

Acton, C., N. Bachman, B. Semenov, E. Wright (2017). A look toward the future in the handling of space science mission geometry. *Planetary and Space Science,* https://doi.org/10.1016/j.pss.2017.02.013.

Arvidson, R., and S. Dueck, 1994, The Planetary Data System, Remote Sensing Reviews, 9, 255-269 https://www.tandfonline.com/doi/pdf/10.1080/02757259409532231.

Besse, S. and 17 others (2018). ESA's Planetary Science Archive: Preserve and present reliable scientific data sets, *Planetary and Space Science*, https://doi.org/10.1016/j.pss.2017.07.013.

Crichton D.J. et al. (2014). Towards an International Planetary Data Standard Based on PDS4. Lunar and Planetary Science Conference Abstracts (LPI Contribution No. 1719), 1815.

Gaddis, L. R., Hare, T.M., and Beyer, R., Summary and Abstracts of the Planetary Data Workshop, June 2012 https://doi.org/10.3133/ofr20141056.

Hughes J.S. et al. (2018). Enabling interoperability in planetary sciences and heliophysics: The case for an information model. Planetary and Space Science 150, 43–49.

McNutt R.L. et al. (2017). Planetary Data System Roadmap Study for 2017–2026, NASA.