
# PDS/EN FY16 Planning

Engineering Node

September 2015

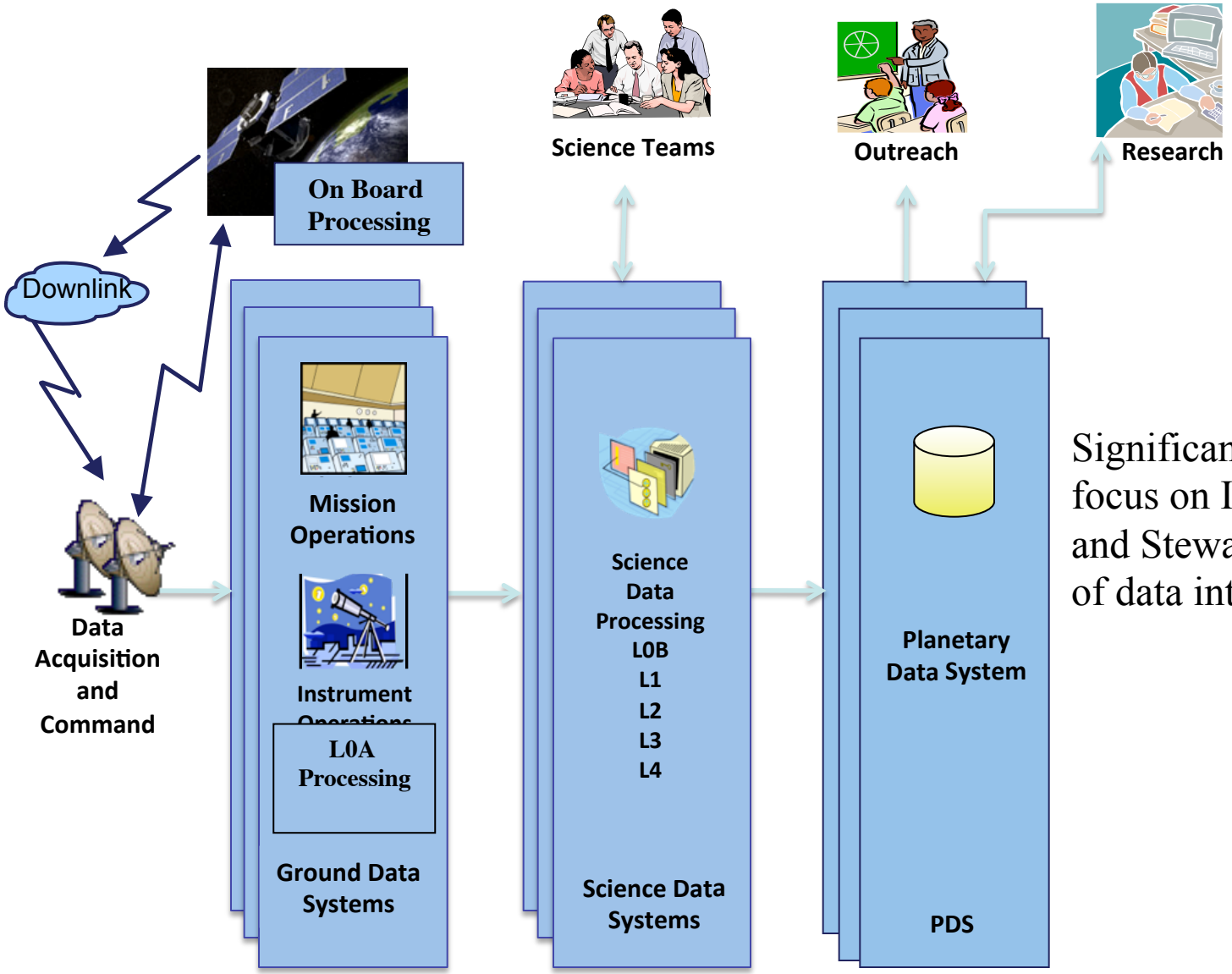
# Key Drivers *

- More Data
- More Complexity (instruments, data)
- More Producer Interfaces
- Greater User Expectations
- Limited Funding
- Creating a system from the federation
- Internationalization
- Increasing IT security threats
- Failover capabilities across the PDS

* Derived from PDS4 Project Plan

"*Support the ongoing effort to evolve the Planetary Data System from an archiving facility to an effective online resource for the NASA and international communities.*" -- Planetary Science Decadal Survey, NRC, 2013-2022

# PDS: Today

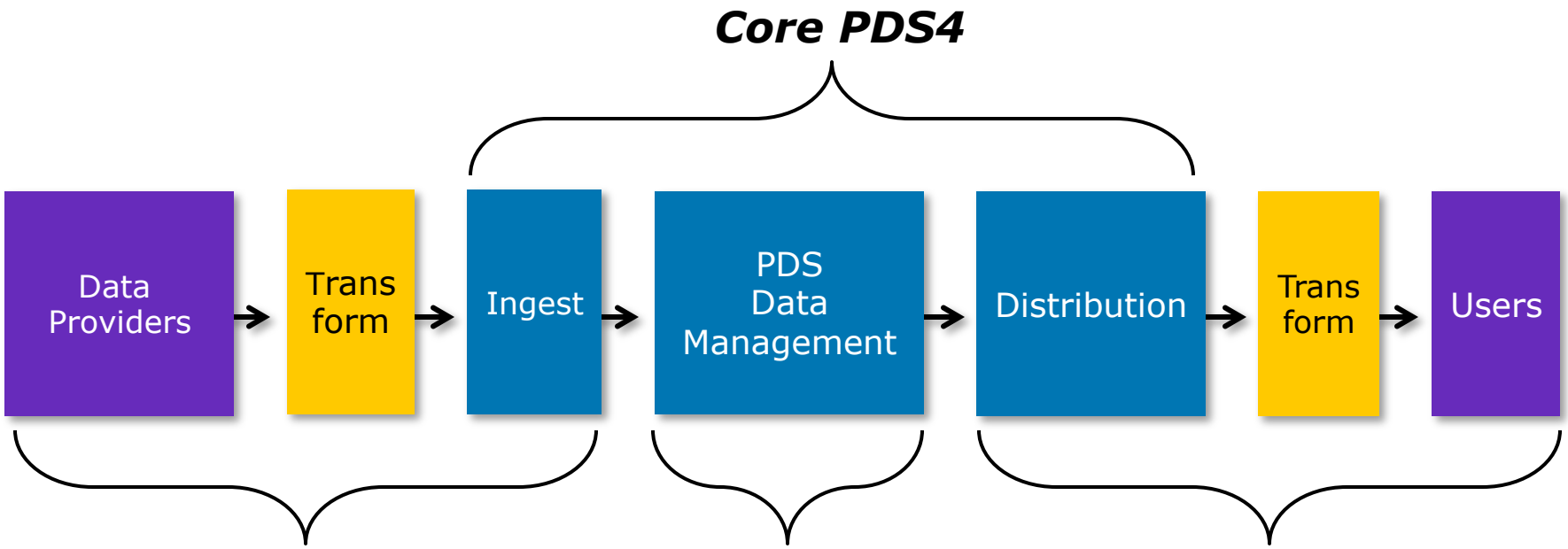



Significant focus on Ingestion and Stewardship of data into/in the PDS

# PDS4 Capabilities: Today

- New information model captured in a modern modeling environment
- Model-driven software services (registry, search, etc)
- Core software tools and services
  - Heavy emphasis on data ingestion and validation
- Full deployment at Engineering Node
  - Integrated with PDS3 and PDS4
- Software installed at all nodes
- Integration underway at nodes
  - LADEE deliveries complete under PDS4
  - MAVEN deliveries under PDS4
- All new U.S. and International planetary missions using PDS4

# Challenge: End-to-End System and Data Integration




1) Improve efficiency and support to deliver high quality science products to PDS

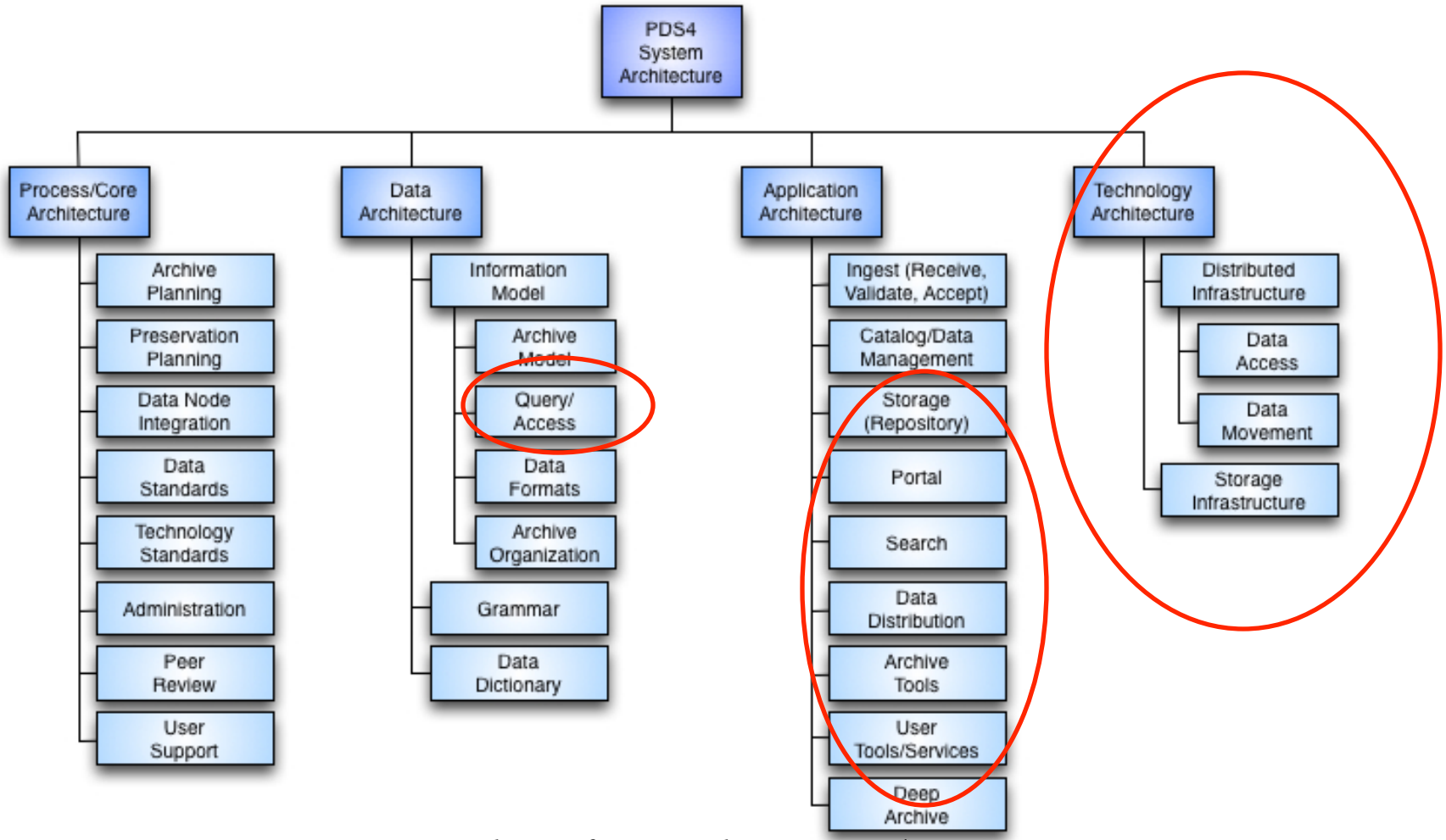
2) Preserve and ensure the stability and integrity of PDS data

3) Improve user support and usability of the data in the archive



Evolving as a system…

# PDS4 System Architecture Decomposition




*2016-2021: Evolving from archive ingest/management to data/node integration, search, distribution and discipline/user tools/support*

# PDS4 Information Architecture/Model Roadmap

| Function | 2010-2015 | 2016-2021 |
|---|---|---|
| **Data Model** | Entire PDS model captured as an explicit model (ontology) defining all aspects including data, missions, instruments, etc | Expanded discipline node and mission models to provide increased capture of mission/science information and provide more tailored user support (search, tools, analysis, etc). Improve integration between software and model. |
| **Data Dictionary** | Captured using a rigorous, well-defined structure based on the ISO/IEC 11179 standard; elements organized into namespaces to allow for international coordination | Online data dictionary registries for mission and user use. |
| **Grammar** | Extensible Markup Language (XML) used to capture PDS metadata; Standard XML tools used | In addition to XML-based support, support for multiple standards for expressing the PDS model (RDF, JSON, etc) to increase use of information model in tools. |

# PDS4 Technical/Software Roadmap

| Function | 2010-2015 | 2016-2021 |
|---|---|---|
| **Ingestion** | Manual process for submission; tools based on PDS4 standards for design/validation | Automated ingestion/submission of data; include increased support for capturing mission information. |
| **Data Management** | Independent data management systems across PDS; initial PDS4 software installed and registration beginning. | Integrated data registries with PDS3 and PDS4 data across the PDS to allow for end-to-end tracking and search; interoperability with international partners. |
| **Storage Management** | Data stored online in independent storage repositories; backup/ failover unique at each node. | Virtualization/commodity storage services to increase integration and reduce cost; PDS-wide disaster recovery and failover in place. |
| **Preservation Planning** | Data maintained in a few simple formats | Transformation services to transform from archive formats to contemporary formats. |
| **Distribution/ Access** | Data distributed in archival format | Enhanced portal for access to data/services/tools; Data distributed in user formats; user services and tools to better facilitate and meet user analysis needs. |

# Benchmark: ESDIS and TCAT

## Action from MSC in Response to TCAT Earth Science Deep Dive

- Assign the Earth Science Division (ESD)/ Earth Science Data Information Systems (ESDIS) to sponsor an *independent* holistic review (with appropriate stakeholders including ESD R&A, ESD Applications, ESDIS, Science community leaders and Mission stakeholders, as well as a senior representative from the NASA CIO community) to **study potential efficiencies and enhanced capabilities** from a variety of perspectives:
  - Based on science discipline
  - Based on optimizing common data operation tasks across DAACs

- **Consider advancing current efforts to achieve efficiencies across DAACs, including cloud computing, open source software , and dataset interoperability**.

# Overall Goals (Next Five Years) – To be worked collaboratively

1. Automate ingestion of data
   - Automated Design; improve label design
   - Automated Ingest (including harvesting)
   - Integrated registry, tracking
2. Capture well-formed, high quality PDS4 data collections
   - Data validated, harvested, registered, tracked
3. Shift from Stewardship to User Services
   - New portal design; PDS website as world-wide site for planetary data and tools
   - Search service deployed at every node
   - Leverage SOA-based architecture for new services, both for Engineering and DN
   - Every product discoverable and accessible
   - Ensure all data products can be visually inspected for peer review
   - Registration and sharing of tools for the planetary science community
4. Establish a more comprehensive virtualization environment
   - Virtualization of 2nd copy
   - Ensure PDS can fail over to secondary storage
5. Upgrade legacy software tools to PDS4
   - Begin decommissioning PDS3 tools
   - Upgrade PDS3 tool to PDS4 ensuring backward compatibility to PDS3

# FY2016 Plans

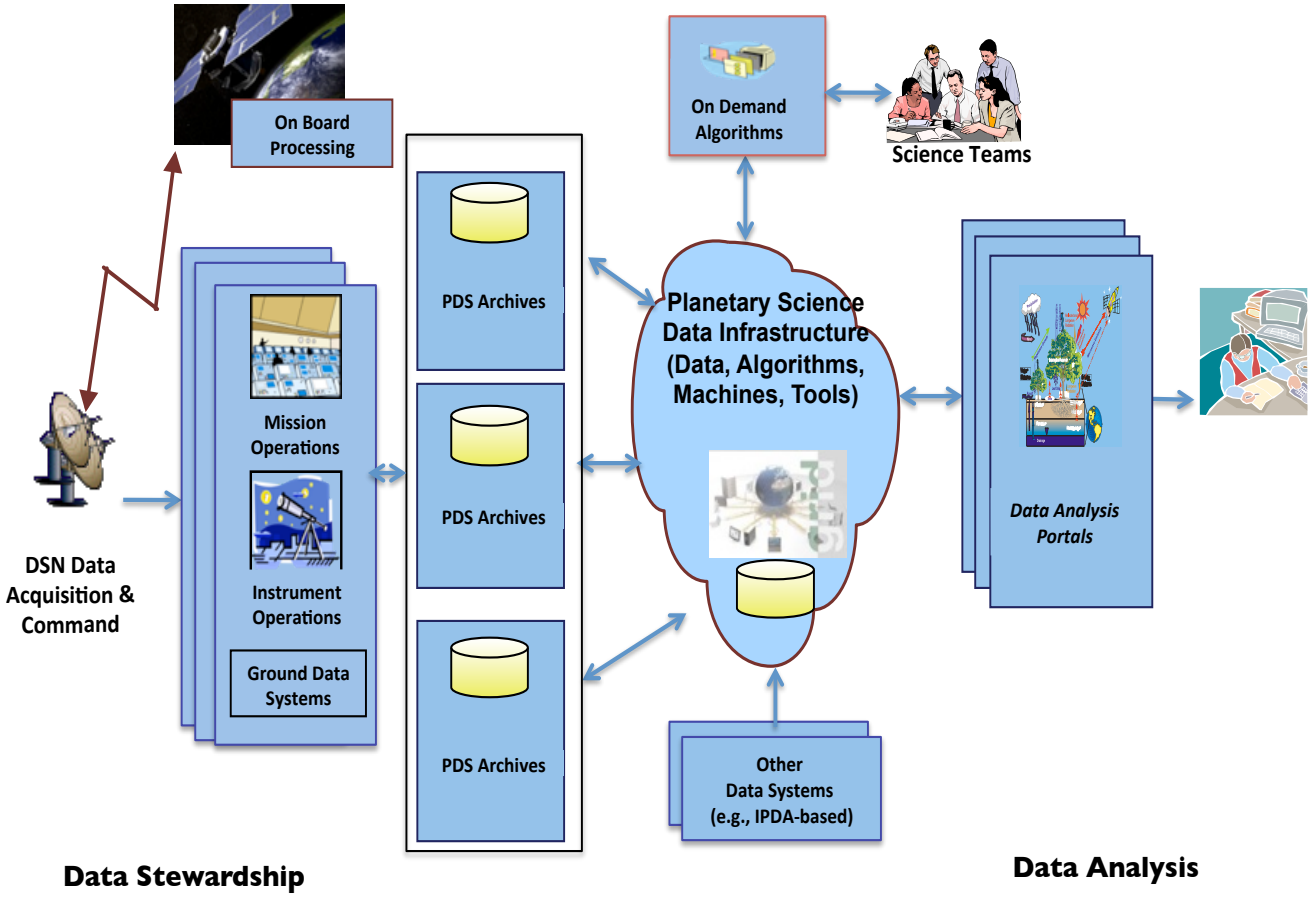
1. Information Model
   - Develop and release new DN model extensions (cartography, geometry, etc)
2. Integrate PDS4 services at the nodes
   - PDS3 registry population; work with nodes to capture schedules for registering PDS3 data
   - Search service integration at the nodes
3. Develop tracking service on top of registry infrastructure
4. Portal: Upgrade to become the world-wide portal for planetary science data and tools integration
5. Virtualization/Storage: Defining requirements for secondary storage service.
6. Expansion of core tools and services (e.g., validation, transformation, search, etc)
7. Conduct two build/releases (Fall '15/Spring '16)
8. Support international expansion of PDS4

# ARC FY16 Support for EN

- Support PDS4 Data Validator development (additional validation checks — metadata, structural, content; in-place editing)

- Incorporate PDS3 validation functionality into PDS4 Data Validator to create a single command-line validation tool

# "An International Platform for Planetary Data Archiving, Management and Research"




"*Support the ongoing effort to evolve the Planetary Data System from an archiving facility to an effective online resource for the NASA and international communities.*" -- Planetary Science Decadal Survey, NRC, 2013-2022

# Summary

- As PDS develops its roadmaps, we are well positioned to transition towards more user services.
  - Consider new working group options to work towards data and system integration
- Continued opportunities to look for automation as we understand PDS4 better
- Opportunities to leverage virtualization, open source technologies, and shared tools at a greater scale
- Considering international partners as part of a larger interoperable system/structure fits well in the PDS4 architecture

# Backup

# Key Req: Archive Standards

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

# Key Req: Tools

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

# Key Req: Cataloging

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

# Key Rqmt: Architecture

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

# Key Req: Search & Retrieval

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

# Key Req: Services

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