PDS 2010
Modernization of PDS

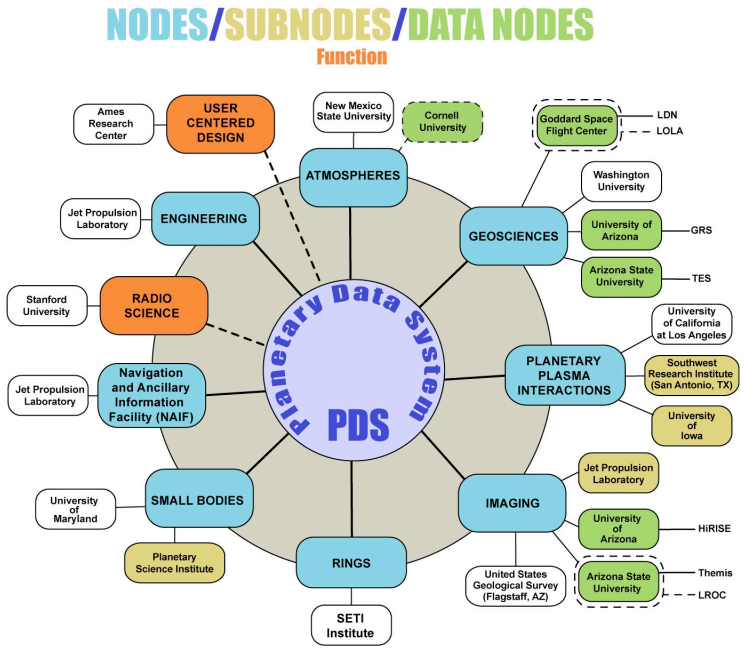
Reta Beebe
Chief Scientist

December 15, 2009

PDS Structure

PDS is composed of service and science discipline nodes and managed by

- Program Executive – Bill Knopf/HQ
- Program Scientist – Michael Kelley/HQ
- Program Manager – Ed Grayzeck/GSFC



In 2005 NASA separated the Central node into The Management Office (GSFC) and an Engineering Node (JPL).

Internally the Management Council –representing each node deals with PDS issues and planned revisions – has monthly telecons and 3 annual meetings – chaired by Reta Beebe who serves as Chief Scientist and coordinator.

PDS Level 1 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
- PDS will collect suitably 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 their usability

Credit: PDS Level 1,2,3 Requirements. August 2006.

Modernization of PDS

- "PDS 2010" is a design effort undertaken to move PDS to a fully online, federated system based on a rigorously defined data model utilizing internationally approved standards
  - The architectural approach allows for better leveraging of modern IT technologies
  - Major effort is in 2010 and 2011

- This effort is driven by several issues which require modernization:
  - rapidly increasing data volume
  - increasing numbers of missions operating simultaneously
  - increasing complexity of instruments and missions
  - more international missions
  - higher user expectations for data manipulation services
  - out-dated standards

Modernization of PDS (cont.)

- General goals of the modernization effort:
  - Improve the efficiency of the data archiving process to lower mission costs
  - Improve the distribution infrastructure and usability of the data to better serve the science community
  - Better address the tension between long-term preservation of the data and usability of the data for current customers
  - Replace aging technology, tools and processes
  - Increase interoperability with other data systems by utilizing international standards for data storage and exchange

Modernization of PDS (cont.)

Specific goals for PDS 2010

- Simplified, but rigorous, archiving standards (PDS V4) that are consistent, easy to learn, and easy to use
- Adaptable tools for designing archives, preparing data, and delivering the results efficiently to PDS
- Online services allowing users to access and transform data quickly from anywhere in the system
- A highly reliable, scalable computing infrastructure that protects the integrity of data, tracks data and archive status, and links the nodes into an integrated data system, providing the best service to both data providers and users

Example Improvements

- Data Products
  - PDS data formats simplified for improved long term preservation and transformation for use in common analysis tools
  - XML, a widely used standard with significant tool support, adopted as the language for PDS labels in PDS 4.0.
  - Creation of syntactically correct labels at the earliest point in the pipeline
  - Enhanced data dictionary structure based on international standards to improve interoperability, provide increased validation capabilities, and provide more finely tuned governance, increasing efficiency across mission, node, and international users

- Software System
  - Updated label-creation and validation tools that rigorously validate against the PDS 4.0 specification
  - Ability to improve searching across the PDS with improved product label definitions

What is PDS 4 / PDS 2010?

- A transition from a 20-year-old collection of standards and tools to a modern system constructed using best practices for data system development.
- Simpler and more rigorously defined formats for science data products.
- Use of XML, a well-supported international standard, for data product labeling, validation, and searching.
- A hierarchy of data dictionaries built to the ISO 11179 standard, designed to increase flexibility, enable complex searches, and make it easier to share data internationally.