

PDS4 Operational Readiness Review PDS4 Project Description, Overall Plan, and Status

Dan Crichton Engineering Node



September 17, 2013





Overview

- Introduction to PDS
- Introduction to PDS4
 - Project, lifecycle, architecture, design
- Build structure and status
- Current deployment and support
- Planned support for LADEE/MAVEN
- Schedule
- Mapping to ToR







PDS Mission and Vision

Mission

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.

Vision

- To gather and preserve the data obtained from exploration of the Solar System by the U.S.
- To facilitate new and exciting discoveries by providing access to and ensuring usability of those data to the worldwide community
- To inspire the public through availability and distribution of the body of knowledge reflected in the PDS data collection

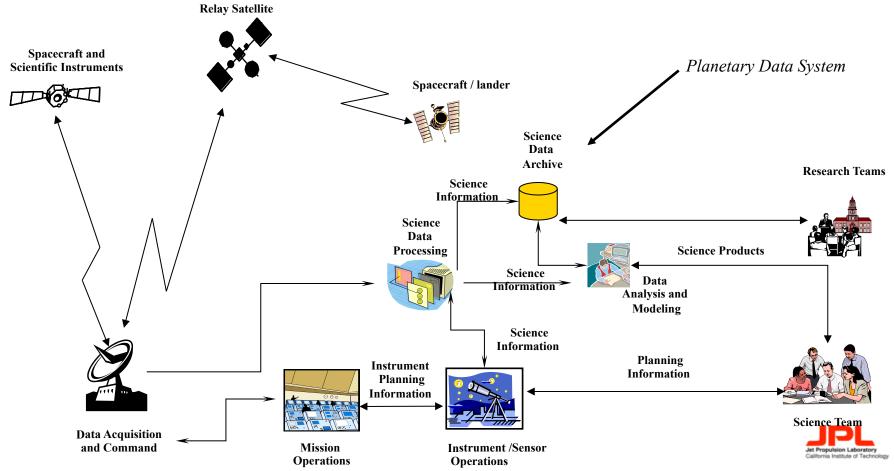
PDS is a <u>federation</u> of distributed discipline and service nodes.







PDS in Context



Credit: CCSDS Reference Architecture for Space Information Management

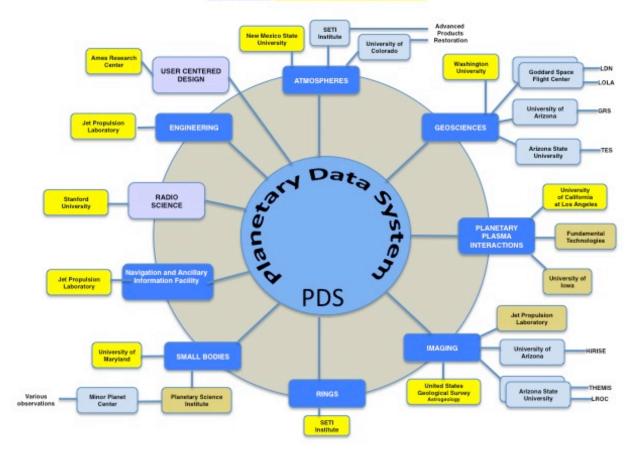


NASA Planetary Data System



NODES/SUBNODES/DATA NODES

Function / Nodes Home Institutions





Timeline of PDS Technical Implementations and Upgrades

- PDS 1 < 1990
 - High-Level Catalog for finding data sets by mission, instrument, spacecraft and target.
 - Archive volumes stored and distributed on tape.
 - The Object Description Language (ODL) is invented for product labeling and capturing catalog information.
- PDS 2 1990
 - CD-ROM becomes the archive and distribution volume of choice.
 - High-Level Catalog simplified by using more text instead of keywords to capture descriptive information.
- PDS 3 1992
 - PDS sets up and maintains a web presence.
 - Movement to online distribution of products (PDS-D). (~2002)
 - On-line mass storage and data bricks replace CD/DVD as archive and distribution media.
- PDS4 > 2010
 - Beginning with prototype/build 1 in 2010
 - Model driven architecture with XML labels
 - Movement to a distributed, service architecture
 - Integrated federation
 - New data standards, data formats and structures
 - International Collaboration







PDS Challenges

- Number and diversity of missions and instruments
 - PDS is currently receiving data from ~110 instruments from 15 active missions as well as concurrently working with missions in development
 - New mission data nodes being added to PDS (LROC, for example)
- Requirements for preservation of data and for usability are sometimes in conflict
- Budget pressures which affect archiving/usability across data providers/missions, PDS and the users
- International archiving and standards coordination
- Increasing volume of data
 - In 2001, the PDS archive was 4 TBs
 - In 2010, the PDS online archive is over 100 TBs
 - In 2013, the PDS online archive is over 500 TBs
- Replacing aging technology, tools, standards and processes







Motivation for PDS4

- The current PDS3 was designed based on an offline system;
 Both the standards and software infrastructure have evolved to support online operations.
- The growth of PDS, both for NASA and non-NASA missions, has stressed the structure and capabilities of the PDS3 standards.
- Software tools, infrastructure, technologies and standards have changed which makes continued maintenance and extension of PDS3 very challenging.
- Ultimately, new software technologies and standards provides an opportunity to greatly improve the operation and usability of the PDS long-term.



- PDS4 is a PDS-wide project to upgrade from PDS version 3 to version 4 to address many of these challenges
- An explicit information architecture
 - All PDS data tied to a common model to improve <u>validation</u> and <u>discovery</u>
 - 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.
- An explicit software/technical architecture
 - Distributed services both within PDS and at international partners
 - Consistent protocols for access to the data and services
 - Deployment of an <u>open source</u> registry infrastructure to track and manage every product in PDS
 - A distributed search infrastructure





PDS4 Planned Mission Support

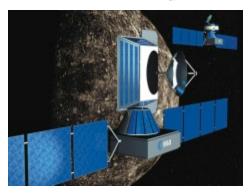




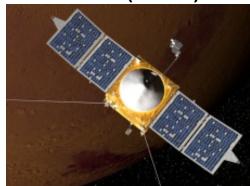
LADEE (NASA)



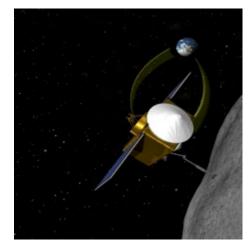
InSight (NASA)



BepiColumbo (ESA/JAXA)



MAVEN (NASA)



Osiris-REx (NASA)



ExoMarso(ESA)

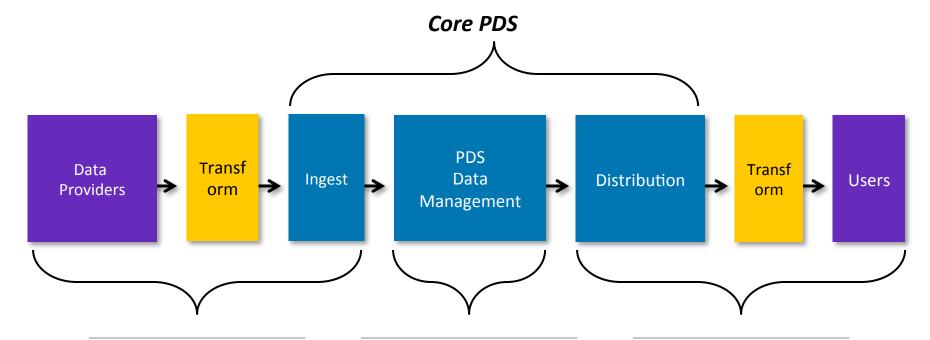


Endorsed by the **International Planetary Data Alliance** in July 2012 – https://planetarydata.org/documents/steering-committee/ipda-endorsements-recommendations-and-actions



Challenge: End-to-End System and Data Integration





Improve efficiency and support to deliver high quality science products to PDS Preserve and ensure the stability and integrity of PDS data Improve user support and usability of the data in the archive







Structured Project Approach

- Phased approach with builds that introduce increasing functionality
- Two key working groups that include members from across PDS
 - DDWG = Data Design WG
 - SDWG = System Design WG
 - Use of several collaboration mechanisms
- Full lifecycle planned out with deliverables for each build (project plan, requirements, design, CM/build, test, release)
 - Reviews at key points in the process







PDS4 Project Leads

- Project Manager: Dan Crichton
- Project Scientist: Reta Beebe
- PDS4 Data Standards Lead: Steve Hughes
- PDS4 System Development Lead: Sean Hardman
- Transition/Operations: Emily Law
- NOTE: Involvement from discipline nodes across the PDS is <u>critical</u>; they are part of the design and development team so they are intimately involved in each step to ensure support for their discipline.

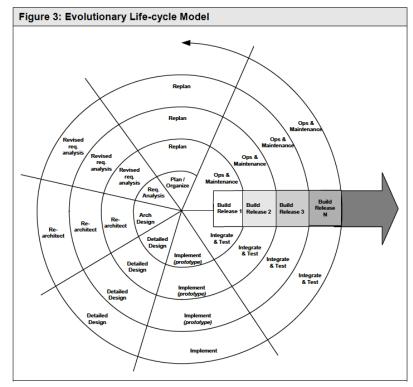
PDS4 Software Development Lifecycle

5.1.3 Evolutionary Development Life-Cycle Model

Table 5-6 summarizes the life cycle defined by the evolutionary development model.

Table 5-6. Summary of Evolutionary Development Life-Cycle Model

Table 5-0	. Summary of Evolutionary Development Life-Cycle Model	
Summary description and discussion	Like the incremental development model, the evolutionary life-cycle model also develops a system in builds, but differs from the incremental model in acknowledging that the user needs are not fully understood and not all requirements can be defined up front. In the evolutionary approach, user needs and system requirements are partially defined up front, then are refined in each succeeding build. The system evolves as the understanding of user needs and the resolution of issues occurs. Prototyping is especially useful in this life-cycle model. (The evolutionary development life-cycle model is sometimes referred to as a spiral development model, but it is not the same as Boehm's spiral model (Reference 11). This model is also sometimes referred to as a prototyping life-cycle model, but it should not be confused with the prototyping technique defined in Section 5.2.1.) This life-cycle model is illustrated in Figure 5–4. Major products and	Revise regardanaly
Advantages	Not all requirements need be known up front Addressing high risk issues (for example, new technologies or unclear requirements) early may reduce risk Like the incremental life-cycle model, interim builds of the product facilitate feeding back changes in subsequent builds Users are actively involved in definition and evaluation of the system Prototyping techniques enable developers to demonstrate functionality to users with minimal of effort Even if time or money runs out, some amount of operational capability is available	Re- architect
Disadvantages	Because not all requirements are well-understood up front, the total effort involved in the project is difficult to estimate early. Therefore, expect accurate estimates only for the next cycle, not for the entire development effort. Less experience on how to manage (progress is difficult to measure) Risk of never-ending evolution (for example, continual "gold plating") May be difficult to manage when cost ceilings or fixed delivery dates are specified Will not be successful without user involvement	J.
Most appropriate when	Requirements or design are not well-defined, not well-understood, or likely to undergo significant changes New or unproved technologies are being introduced System capabilities can be demonstrated for evaluation by users	<u>Ev</u> ND MAVEN



JPL-D 76772 Rev 1
NASA Software Management
Guidebook - NASA-GB-001
Evaluation by users as early as possible!

14

Project Lifecycle Thru Build 1 & 2

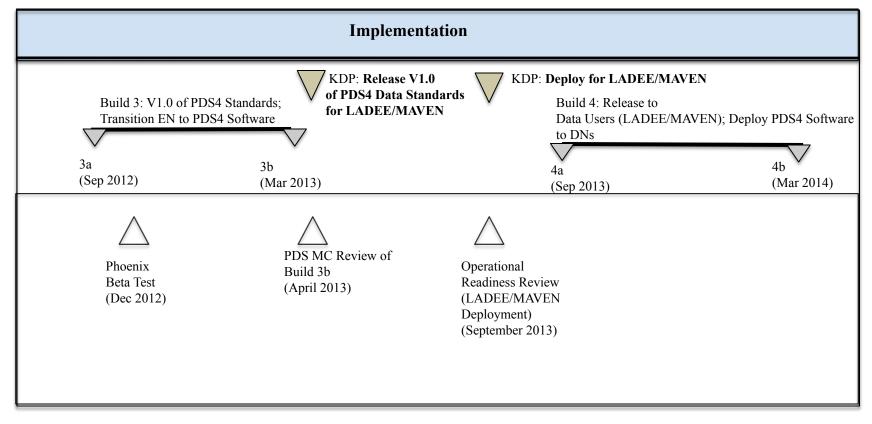
Project Lifecycle	Pre- Formulation	Formulation	Implementation	
Events	RDP: Study Begin Study/ Study Concepts Project	\triangle	Build 1: Prototype build 1a 1b 1c 1d (Oct 2010) (Aug 2011)	KDP: Beta Release for LADEE/ MAVEN re for label design 2b (Mar 2012)
Project Reviews	PDS MC Concept Review (Dec 2007)	PDS MC PDS MC Impl Arch Review Review (July 2008) (Nov 2008)	PDS External System System Design PDS External System Design Review II	ORR (Start Label Design) (LADEE/ MAVEN) (November 2011)

Note: These will be discussed in detail in a later presentation



Project Lifecycle Builds 3 and Build 4









Key Project Documentation

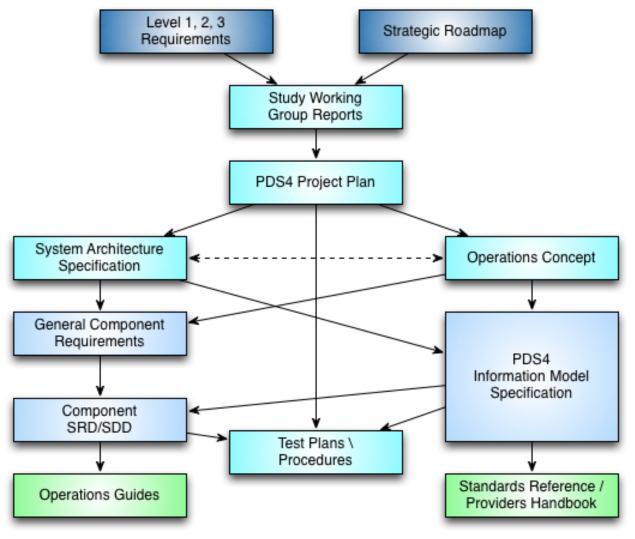
- Requirements
 - PDS Level 1, 2, 3
- Project Plan
 - Contains management approach and implementation plan
- Multiple Concept Papers
 - Architecture, user services and data design
- System Architecture
 - Defines data and software architecture
- Operations Concept
 - Interactions of PDS across the mission phases and from ingestion thru to distribution
- System Design Specifications for services and tools
- In addition, PDS maintains policies, requirements, standards, and schedule information online at the Management Council and Engineering nodes







Document Tree







Summary of Progress to Date

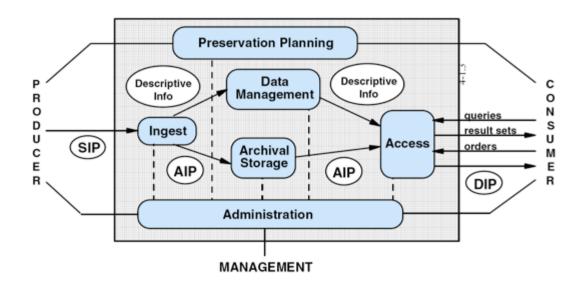


- Requirements in place (approved by MC 3/2010)
- PDS-wide Architecture defined
- Major reviews conducted both internal and external (data providers, users, engineers)
- Initial PDS products defined using maturing PDS4 specification
 - LADEE, MAVEN teams developing PDS4 labels per ORR and decision in Nov 2011/Jan 2012
- System builds grouped by purpose: build 1(a,b,c,d), 2 (a,b,c), 3(a,b)
 - Used to establish a rigorous release process
 - Includes deployment of software, data standards, documents and system integration testing
 - As stability has increased, we have moved to 6 month builds
- Operational capabilities deployed
 - Registry and Harvest infrastructure in place at EN
 - PDS3 Central catalog migrated to PDS4 registry; High level search migrated
 - PDS3 data now being ingested into PDS4 system at EN
 - PDS4 Standards Released to V1.0 to prepare for LADEE/MAVEN; Change Control Board enacted for V1.0 and beyond.
- Beta test conducted by Atmospheres to get user input on PDS4 concepts (XML, Bundles, etc)
- LADEE PDS4 Peer Reviews conducted
- IPDA endorsement and plans to move to PDS4
 - International implementation by PSA for Bepi Colombo underway
 - Deployed international search using PDS4 at the data set level





- Ingestion
- Data
 Management
- Storage Management
- Administration
- Preservation Planning
- Distribution/ Access



Reference Model for Open Archive Information System, CCSDS 650.0-B-1, January 2002



PDS4 Technical Implementation PDS Differences

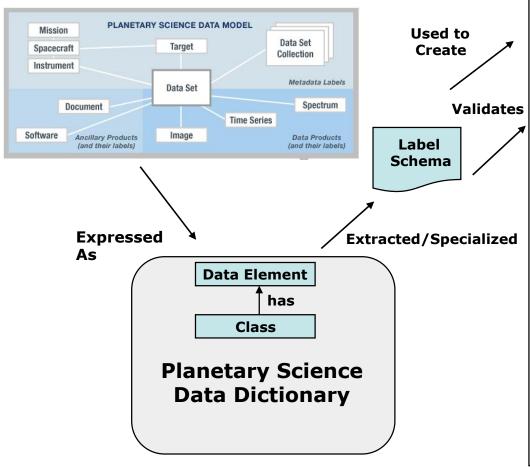
Function	PDS 3 Implementation	PDS 4 Implementation
Ingestion	Manual process for submission; tools based on PDS internal standards	Automated ingestion; XML-based tools for design, validation and submission
Data Management	Independent data management systems across PDS	Integrated data registries across the PDS to allow for end-to-end tracking and search; interoperability with international partners
Storage Management	All data being migrated online	All data stored online
Preservation Planning	Missions deliver data formatted for the archive	Data maintained in a few simple formats that allow for transformation and long-term use
Distribution/ Access	Data distributed in archival format	Data distributed in user formats; user services to better support analysis

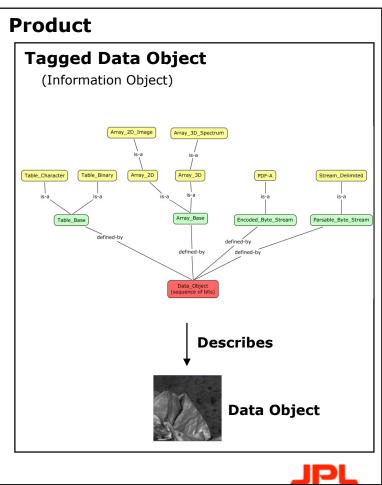


Data Design Approach



Information Model



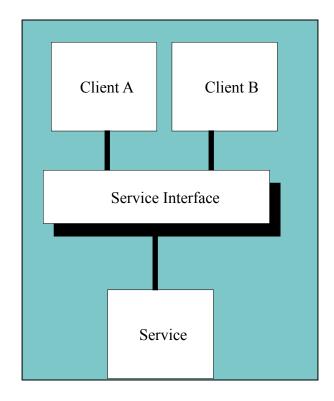




System Design Approach



- Based on a distributed information services architecture (aka SOA-style)
 - Allow for common and node specific network-based (e.g., REST) services.
 - Allow for integrating with other systems through IPDA standards.
- System includes services, tools and applications
- Use of online registries across the PDS to track and share information about PDS holdings
- Implement distributed services that bring PDS forward into the online era of running a national data system
 - With good data standards, they become critical to ultimately improving the usability of PDS
 - Support on-demand transformation to/from PDS





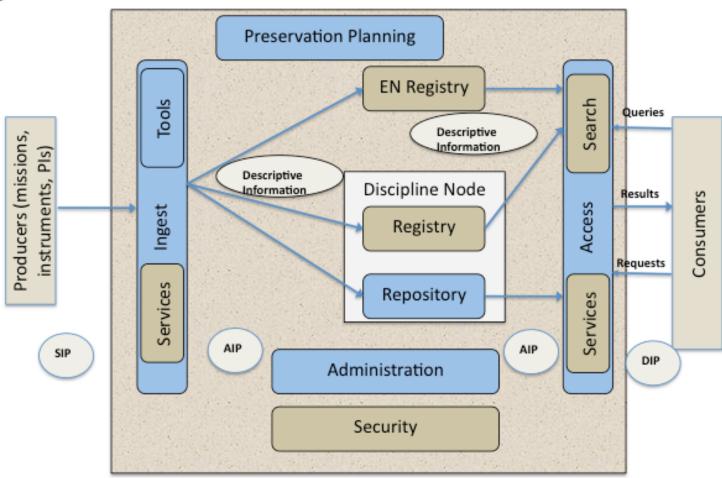
PDS4 Data Standards Differences System System Standards Differences System Syst

Function	PDS 3 Implementation	PDS 4 Implementation
Data Model	High level information model; ad hoc model for each data set/product	Entire PDS model captured as an explicit model (ontology) defining all aspects including data, missions, instruments, etc
Data Dictionary	Based on a PDS internal structure	Captured using a rigorous, well-defined structure based on the ISO/IEC 11179 standard; elements organized into namespaces to allow for international coordination
Grammar	Object Description Language (ODL) used to capture metadata and annotate data sets, products, and catalog files	Extensible Markup Language (XML) used to capture PDS metadata; Standard XML tools used; <i>separation of the storage and display formats</i> .



PDS to OAIS Mapping





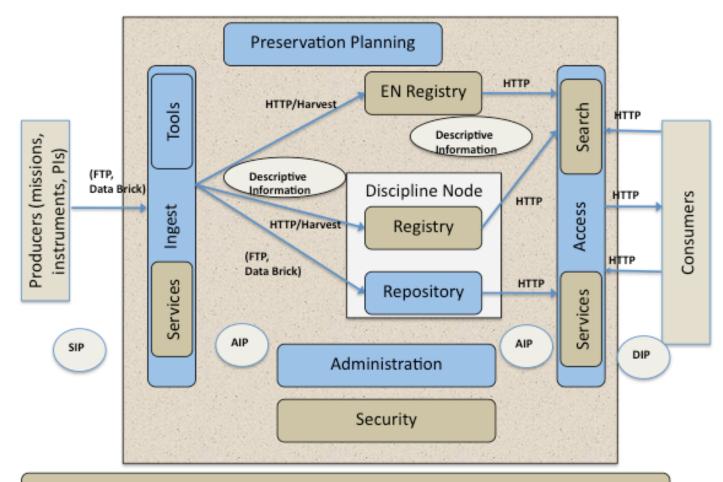
Governing Information Model and Data Standards





PDS to OAIS Mapping





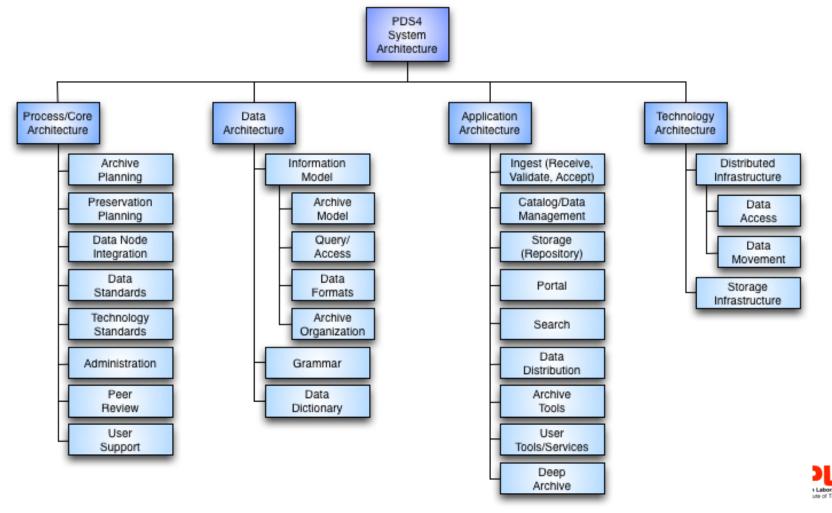
Governing Information Model and Data Standards

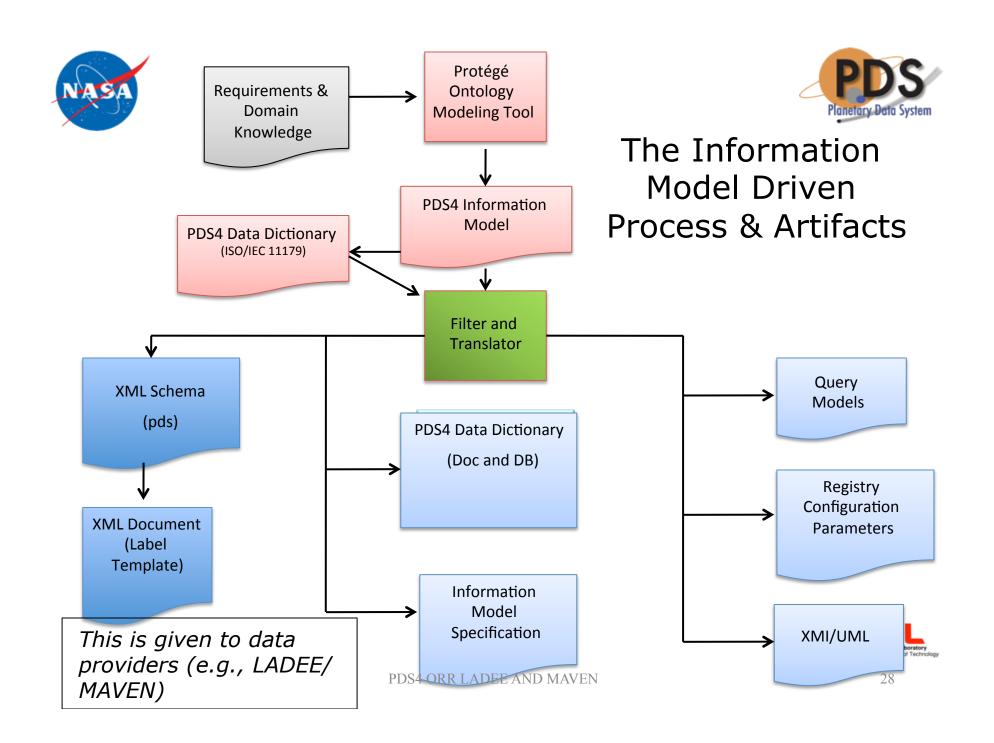




PDS4 System Architecture Decomposition



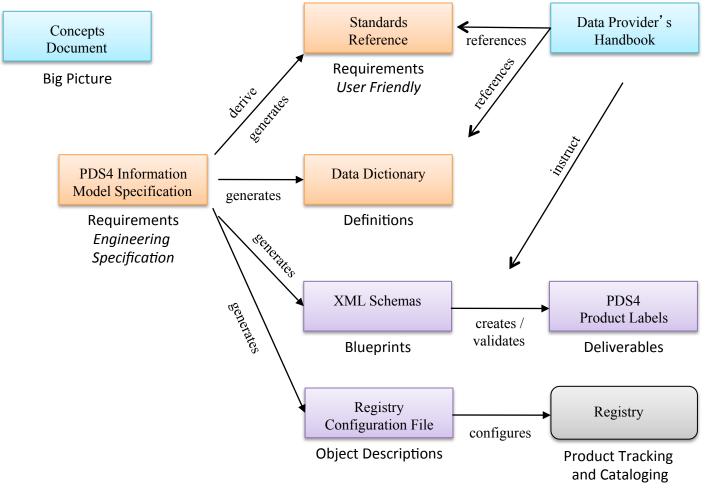






PDS4 Documents & Artifacts











System Builds

- PDS4 uses system builds to bring together the software and the information model
 - Established very early in the project to organize releases
 - Provides a predictable structure to bring the teams together
 - Provides incremental functionality
 - Allows for feedback both inside and outside PDS
 - Allows for adoption by the DNs which have varying needs over time
- Each build provides a full lifecycle to capture, CM, integrate, test and deploy the release





PDS4: Build Structure



Phase	Purpose	Release	Date
Build 1 Prototype/ Test	 Early formalization of the data standards Early formalization of software Integration between software and data standards Stakeholder input 	 PDS4 beta info model, standards reference, data dictionary, schemas baseline Early releases of Harvest, Registry and Security services First set of process, documentation and tutorial 	Oct 2010 Feb 2010 May 201 Aug 2011
Build 2 Begin Label Design for LADEE/ MAVEN	 Support LADEE, MAVEN Label Design and Planning Begin pilot deployment of PDS4 at Engineering Node and transition 	 PDS4 beta info model, standards reference, data dictionary, schemas baseline Release of Harvest, Registry, Report and Security services Validation and catalog ingest tools Updated documentation 	Sept 2011 Feb 2012 June 2012
Build 3 V1.0 of PDS Data Standards for LADEE/ MAVEN	 Transition entire PDS3 catalog to PDS4 Baseline PDS4 standards as version 1.0 Support validation of PDS4 bundles Support ingestion and PDS4 data into the PDS4 registry Support search and access to PDS3 data sets and PDS4 bundles Support LADEE/MAVEN data ingestion 	 Stable release of data standards Incremental releases of validation and PD3 catalog ingest tools Incremental releases of harvest, registry, report and security services Deployment of Registry Services at EN Deployment of the PDS4 search service at EN 	Sept 2012 March 2013
Build 4 User Services	 Support PDS4 data distribution services for LADEE, MAVAEN Support PDS4 data transformation PDS4 user tools PDS4 ORR LA 	• Incremental release of PDS4 Standards • Distribution services for initial PDS4 bundles • Transformation services for initial PDS4 Dbandles MAVEN	Sept 2013 March 2014





Build 3b Deliverables*

- Software System
- Registry Service
- Harvest Tool
- Validate Tool *
- Security Service
- Report Service
- Search Service
- Catalog Ingest Tool
- Upgraded portal search and page views to support PDS4

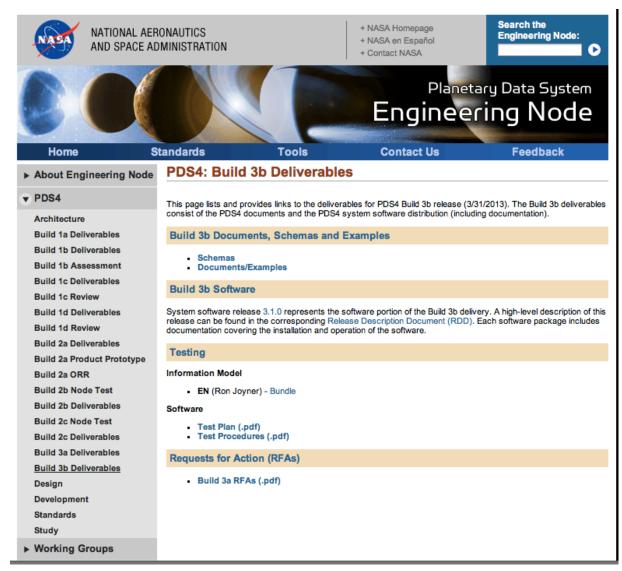
- Data Standards*
- Information Model
- XML Schemas
- Data Dictionary
- Concepts Document
- Standards Reference
- Data Providers Handbook
- PDS4 Example Products



^{*} Posted to http://pds.nasa.gov/pds4



Build 3b Screenshot







Key PDS4 Components for LADEE/MAVEN



- Software System
- Registry Service
- Harvest Tool
- Validate Tool *
- Search Service
- PDS Portal

- Data Standards*
- Information Model
- XML Schemas
- Data Dictionary
- Concepts Document
- Standards Reference
- Data Providers Handbook
- PDS4 Example Products









Build Testing

- Test plans provide regression testing for each build
 - Test cases traced to requirements
 - Example data product types are used as input
- Engineering performs regression tests using example bundles and data products
 - Validate functional requirements L3 L5
- Engineering generates a test report
- For build 3b, the PDS MC voted to accept the data standards/IM as V1.0 to support LADEE/MAVEN
- For build 3b, PPI and ATMOS have participated to validate MAVEN and LADEE needs
- Testing and results will be discussed tomorrow





Deployed Capabilities for Key PDS Functional Areas



- Design and Generate PDS4 Products
 - COTS tools tested and in use to support PDS4 Schema development (build 2)
 - DN tools for generating PDS4 products developed at multiple nodes (build 2)
 - Information Model V1 released (<u>build 3b</u>); CCB established.
- Validation
 - Validate tool developed for PDS4 label validation (build 2)
- Harvest/Registration
 - PDS3 central catalog migrated to registry (build 3a)
 - CI tool in place to register PDS3 catalog data (done, e.g., MSL) in a PDS4 registry (build 3a)
 - Harvest in place to register PDS4 bundles and resources in PDS4 registry (build 3a)
 - Registration of PDS4 web resources (build 3a)
 - Registration of PDS4 bundles (build 3b)
- Search/Access
 - Deployment of the PDS4 search service at EN (build 3a)
 - Generation of a PDS4 search index for PDS3 data (build 3a)
 - High level search of PDS3 data sets, PDS4 bundles, web resources, IPDA (build 3b)
- Distribution/user tools (<u>build 4</u>)
 - Deploy build 4a (V1.1 of standards) for LADEE/MAVEN production use (build 4a)
 - Support product level search for LADEE/MAVEN (build 4a)
 - Begin planning support for O-Rex and Insight (build 4b)





Transition from PDS3 to PDS4



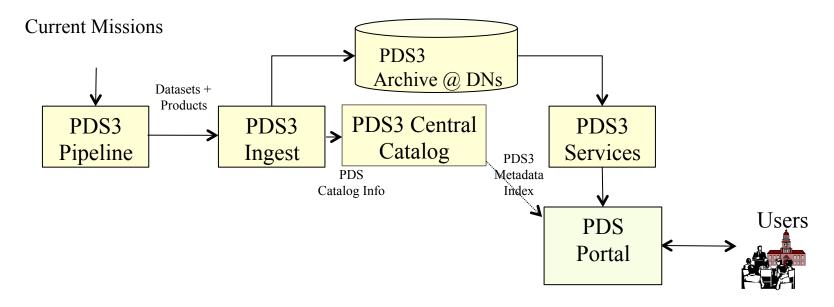
- PDS designed PDS4 to be able to access both PDS3 and PDS4 archives through a single system given that
 - PDS is required to continue to support PDS3 data deliveries from active missions (e.g., Cassini, MSL, etc)
 - Discipline Node adoption to PDS4 will occur independently
 - Desire to support a single, operational system
- The architecture of PDS4 has allowed the Engineering Node to do this by enabling "registration" of PDS3 and PDS4 data
 - PDS3 catalog information explicitly included in the model as a PDS4 product type
 - To date, the PDS3 "catalog" information has been registered as PDS4 products supporting search/access to the distributed PDS3 archives
- This will be discussed in detail tomorrow





PDS3 Implementation

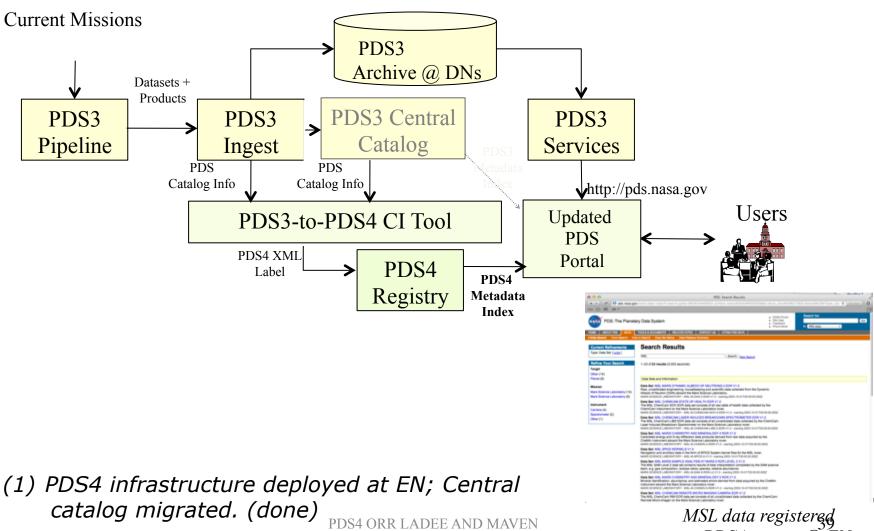






Today



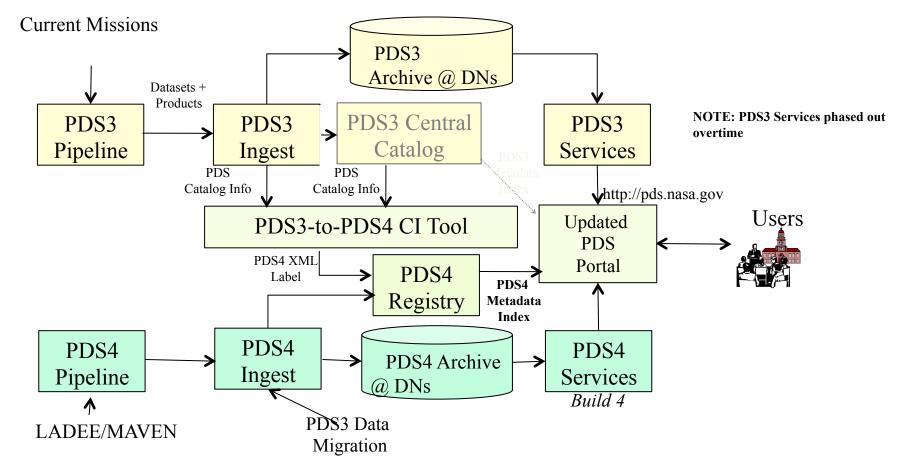


MSL data registered in PDS4 system @ EN



Planned Support for LADEE/MAVEN





- (1) PDS4 infrastructure deployed at EN; Central catalog migrated. (done)
- (2) Working towards acceptance of LADEE/MAVEN PDS4 mission data



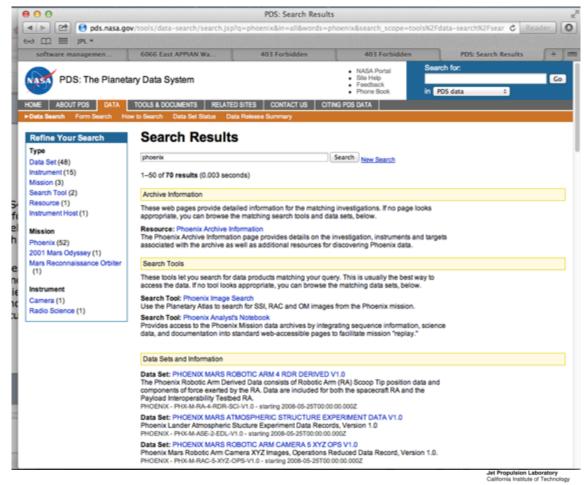




Beta Testing

- PDS4 development has performed increasing levels of beta testing with users since build 1
- Latest beta test by Atmospheres included review of PDS4 labels, bundles and documentation

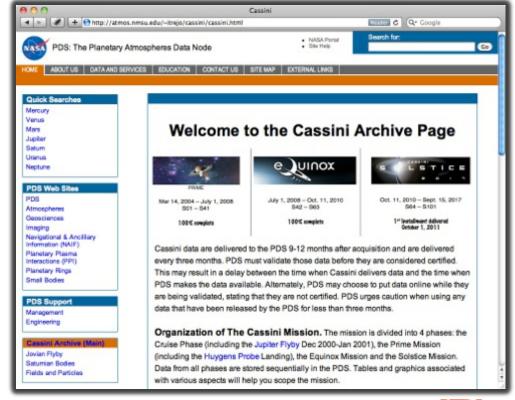
Search Service Deployed for Build 3a Includes links to Archive Resources





PDS4 Archive Support Pages

- A new concept in PDS4
 which allows websites to be
 registered, searched,
 accessed and used by
 novice users to get to
 archival data
- Cassini Archive Support page has been received well
- LADEE and MAVEN pages already developed and sent to teams to review
- Already deployed for Cassini





LADEE and MAVEN Support



- Atmospheres and PPI Nodes will discuss their commitments, timeline, and support using PDS4 in later presentations.
- To date,
 - Nodes have worked directly with these missions to support use of PDS4
 - PDS4 label design began after build 2b with concurrence from the PDS MC in Winter 2011. Templates provided to the missions.
 - V1.0 in place with CCB to support LADEE and MAVEN deployment
 - A few change requests for MAVEN will be included in build 4a to allow conforming CDF files to be captured in PDS4 as an archival product.
 - Additional improvements in metadata to support better faceted search will be in V1.1
 - PDS4 product types to be used tested to support design, harvest, registration, and search/distribution.
 - LADEE and MAVEN archive pages developed.
 - Build 3b software infrastructure already deployed at Engineering, Atmos, and PPI. Used operationally at EN to support all PDS data.



Schedule



- Build 4a delivery September 30, 2013
- Build 4a testing Sept 30, 2013 Oct 31, 2013
- Build 4a release October 31, 2013
- Deployment of build 4a to Atmospheres and PPI November/ December 2013
 - Upgrade registry software at nodes (as needed)
 - Production deployment of the search service
- LADEE Data Delivery & Distribution March 2014
- LADEE Final Delivery June 2014
- MAVEN Data Delivery & Distribution April 2015
- MAVEN Deliveries every 3 month until EOM

Build 4a will provide the baseline with V1.1 of the PDS4 model, however, additional capabilities can be provided in future builds as needed.





Risks Raised in ORR-1

- PDS3 Operations ensure we can continue supporting PDS3 operations with little to no disruption (ORR-1)
 - The transition from a PDS3 infrastructure to a PDS4 infrastructure supporting PDS3 data has mitigated this risk.
- PDS4-Mission Exposure using PDS4 for label design (ORR-1)
 - The limited release of PDS4 is allowing for it to be validated incrementally.
 - The delivery of templates to the missions has been an enormous help in lowering the learning curve and constraining the labels.
 - The LADEE and MAVEN experience has provided excellent input for future missions.





ORR Risks for this Deployment



- PDS4 Formatted CDF Files for MAVEN
 - Description: PDS is planning to accept a PDS4 compliant CDF file for archive. This was negotiated after build 3b and V1.0 and confirmed by the MAVEN mission for their ORR in August.
 - Risk: PDS has not accepted CDF types files into the archive since it is not an international standard.
 - Mitigation
 - PPI has developed a mapping from CDF to PDS4.
 - The PDS4 CCB has approved very minor changes to V1.0 to accommodate the approach. To be released in V1.1.
 - Should CDF present a problem, the data can be converted and archived as a PDS4 table structure.
- Changes to PDS4 Data Standards
 - Description: Changes and Improvements to PDS4 Standards can impact missions if they are required to make changes.
 - Risk: Changes to PDS4 Standards that impact LADEE and/or MAVEN
 - Mitigation:
 - When selecting the first PDS4 missions, Atmos and PPI agreed to upgrading data products, if required.
 - The CCB is intended to ensure that stability is in place and to limit changes which affect LADEE and MAVEN.
 - With V1.0, the number and scope of changes has dropped drastically, but improvements will be made. PDS is testing an upgrade of summary/search metadata that is scheduled to be release for V1.1 which PPI and Atmos





ORR ToR/PDS4 Mapping

- Entrance Information*
 - PDS4 Standards Documents (Standards Reference, Data Providers Handbook, Concepts Document)
 - PDS4 Examples
 - PDS4 Tool
 - Schemas
 - Schedule of deliveries and deployment



^{*} See http://pds.nasa.gov/pds4/orr0913/index.cfm



ORR ToR/PDS4 Mapping



Success Criteria

- All major issues and/or liens/actions identified during the review have been resolved and/or plans established to resolve them.
 - We will work with the chair to ensure RFAs are addressed.
- PDS4 documents are complete and usable.
 - The documents have continued to be aligned with each build and have gone through significant review inside and outside PDS. They have been developed in parallel as early support was provided to the missions.
- PDS4 examples are validated and registered.
 - The test cases and results which show the validation of V1.0 data products, the harvesting of those products, and the registration, will be shown to the board.
- PDS4 deliveries tested by lead nodes for LADEE and MAVEN.
 - Atmospheres and PPI have tested both simulated data products and software. Atmospheres has used migrated data from Phoenix. PPI has used data from MGS and ARTEMIS.
- Core tools, both off the shelf and custom developed, are in place to support design and validation.
 - Teams have used Oxygen for designing labels. Templates have been delivered to LADEE and MAVEN. The Validate Tool provides validation support for V1.0 (and V1.1 with the build 4a release)



Resources



- http://pds.nasa.gov/pds4 (PDS4 site)
- http://pds-engineering.jpl.nasa.gov/ index.cfm?pid=145&cid=190 (CCB Information)
- http://pds-engineering.jpl.nasa.gov/ index.cfm?pid=145&cid=189 (key engineering information: project plan, requirements, architecture, design, reviews, links to builds, etc)

