
# PDS 2010: Overview

**4th IPDA Meeting, Rome**

**Dan Crichton**

**July 2009**



# Caveat!

- This is “hot” off the design press
- IPDA is getting a first glimpse of the concept, tradeoffs and recommendations that are being formulated for the MC
- We solicit an encourage your feedback, questions and involvement



# What is PDS 2010

- A major upgrade to PDS to become an online, “service oriented” data system
- Improvement of data standards to PDS v4.0 to ensure consistency and simple adoption
  - Improved management of data formats
  - Better grammar support
- Overhaul of the system to improve long term operations and manage cost
- Improved tracking from data provider to deep archive
- A framework for building online, user services



# PDS 2010 Goals

- Specific goals for PDS 2010 *
  - Simplified, but rigorous, archiving standards (PDS4) that are consistent, easy to learn, and easy to use
  - Adaptable tools for designing archives, preparing data, and delivering the results efficiently to PDS
  - On-line 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, links the nodes into an integrated data system, and provides the best service to both data providers and users

* PDS 2010 Executive Summary, July 2008

# Summary of Progress

- PDS put together and developed an architecture for PDS 2010 that was presented to PDS in November 2008
- After November, PDS formed two key design working groups to address
  - System design
  - Data design
- The preliminary designs were recently discussed at a technical session held at JPL on June 9-11, 2009
- The technical progress will be reported to the MC with a plan to move to implementation

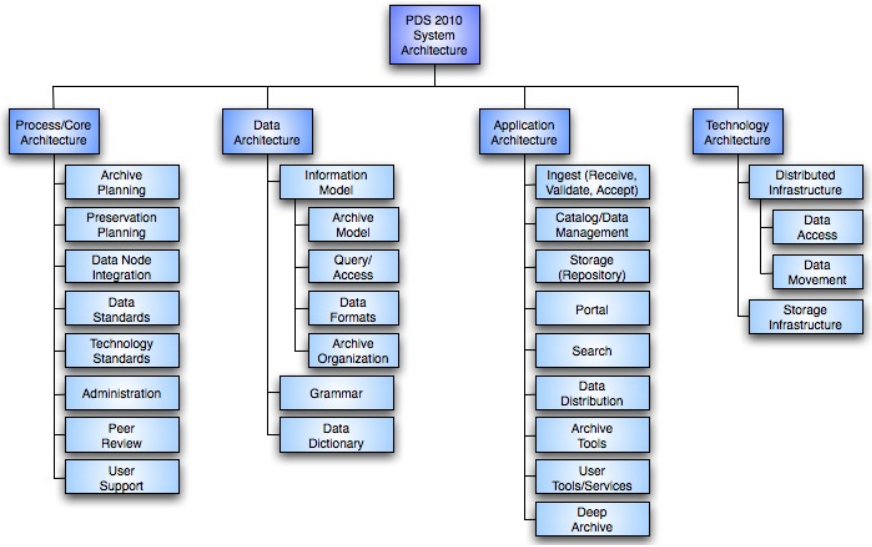


# Key Architecture Decisions (as presented in Nov to MC)

- Move towards a software services model which allows for packaging and reuse of software as services over the network
- Develop a PDS4 model that encompasses both description of "data objects" and definition of data structure
- Identify opportunities to leverage newer standards, where appropriate
- The PDS 2010 design designs discussed at the tech session included discussion of tradeoffs and impacts

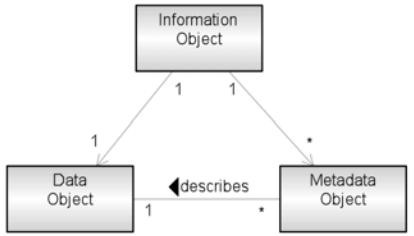


# Elements of PDS 2010




# Data Architecture Concepts

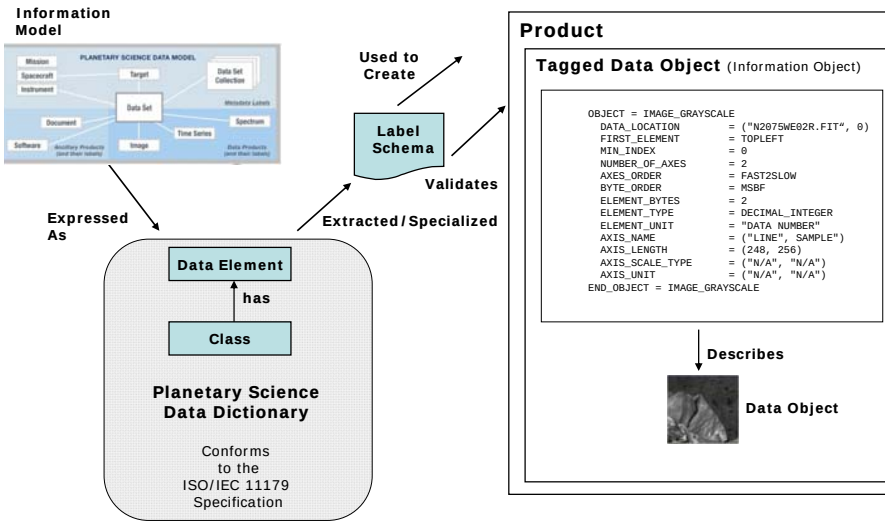
- A PDS data product is defined as having
  - A data object (the actual science data bits)
  - Metadata object
- In PDS 3, PDS does not have standard formats for data objects
- In PDS 4, PDS will have structure




Information Object *

* CCSDS Reference Architecture for Space Information Management (RASIM), 2006

# Data Architecture Concepts




# Technical Architecture Concepts

- What do we mean by "software services"?
  - Software services are online, callable services that perform a function and return a result. They can be invoked over the network and can be shared.
- One of the primary focuses will be defining the "connectors" for the services
  - These will encompass some industry standard interfaces and some PDS-specific ones

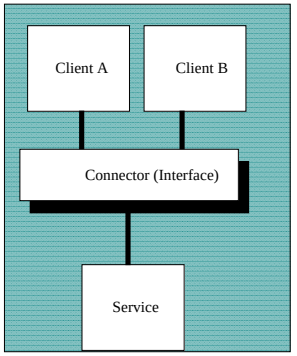


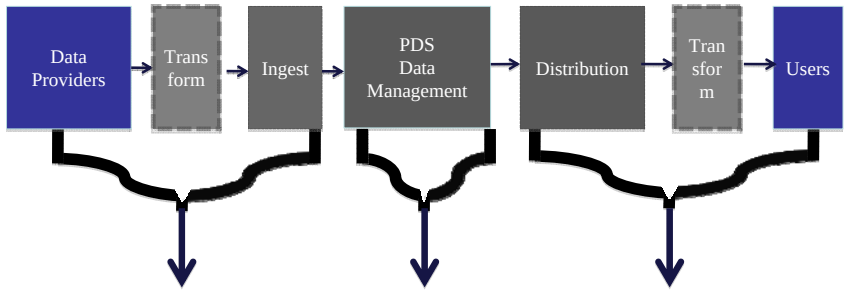

Figure – Component Connector Model (C2).



# Focusing the Design

- Delivery to PDS
  - Provide data standards and software tools and services to improve the efficiency of delivery of high quality products to PDS
- Management within PDS
  - Preserve and ensure the stability and integrity of the PDS data
- Distribution from PDS
  - Provide software services, tools and standards to improve the usability of PDS and the data in the archive

# Level 0 Conceptual Flow

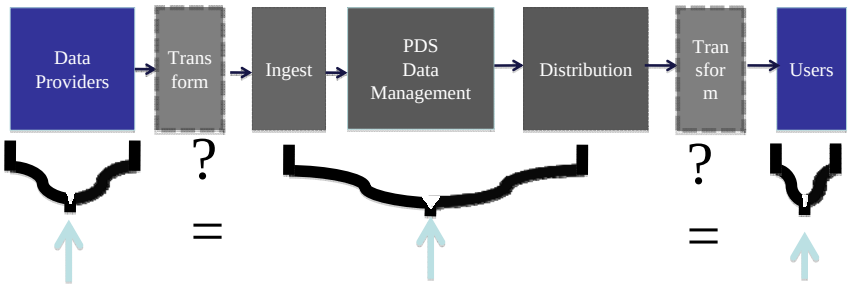



**Goal**: Improve efficiency and support to deliver high quality science products to PDS

**Goal**: Preserve and ensure the stability and integrity of PDS data

**Goal**: Improve user support and usability of the data in the archive

# Level 0 Conceptual Flow (Data Model View)

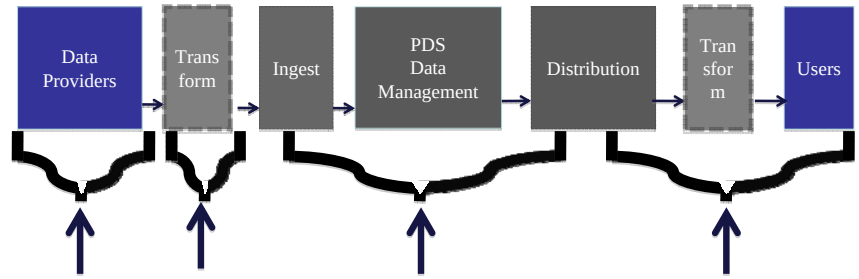



**Data Provider Model (DF1)**: Data formatted based on data producer pipeline

**Archive Model (DF2):** Data formatted for long-term archive

**User Model (DF3):** Data formatted for user use

# Level 0 Conceptual Flow (Software Mapping)




**Tools**: Generation, Design, Validation

**Tools**: *toPDS*

**Tools**: Validation, Ingest

**Services**: Inventory Management

**Services**: Search, Data Discovery, Data Access, *fromPDS*, DN services

# Two Key Design Considerations

- Ensure the design of PDS allows for phased improvements over time to either/both ingestion and data distribution (e.g., delivery of data in user formats)

  Rationale: PDS needs to be able to deliver capabilities in phases in order to manage both the operational and budgetary demands and limitations. This will tie all the way to the transition and impact discussion.

- Provide clear boundaries between “archive”, “user”, “data providers” for data, support and services

  Rationale: Drives cleaner system designs and also allows us to understand the system in context

# Key Trade offs

- PDS4 Data Design and Standards Mgmt
  - PDS data/information model and new structures
  - PDS grammar (XML vs ODL/PVL)
  - Data Dictionary (to ISO 11179)
- PDS System Design
  - PDS Distributed Infrastructure (Registry, Data Dictionary, etc)
  - Format Transformation Tools and Services (to/from PDS)
  - Federated Registries (including Product-level)

# Design Decisions (1)

| | Resources | Training | Usability | Efficiency |
|---|---|---|---|---|
| PDS4 Model | Short term impact on developing/migrating model and dictionary; long term costs to maintain standards decrease with more stability in the model | Data engineers, providers will need to be trained in order to develop PDS4 compliant data products | Usability for data providers and users improved if format transformation is in place | Improved archiving efficiency; better support for design, submission and use of PDS data |
| XML | Short term impact to re-write software for XML; longer term costs decrease by using open source libraries | Data engineers, providers and software developers will need to be trained in XML | XML data can be transformed for viewing purposes, etc | Will streamline software development of tools and services (e.g., data dictionary service) |
| ODL+/PVL | Short term impact to upgrade software for ODL+/PVL; longer term, costs will be incurred to maintain grammar language and software | Some re-training will be required | Will require development of specific ODL/PVL tools | Long term, ODL/PVL is a less efficient approach than XML due to its limited adoption, support and obsolescence risk |



# Design Decisions (2)

| | Resources | Training | Usability | Efficiency |
|---|---|---|---|---|
| Distributed Infrastructure | Major short-term impact will be on system development; Improved efficiency, tracking, access to data realized | Primary training issue is in developing software which can access online services over the network | Better support for design, validation and submission to PDS; more unified PDS-wide access to data; International access and support will be enabled. | Improved efficiency by reducing the number of versions of ad hoc tools and information |
| Format Transformation Tools and Services | Primary impact will be on developing tools to convert to/from PDS. | Primary training issue is in developing software that can perform transformation to/from PDS4 | Improved usability by allowing users to access data in user preferred formats. | Improved efficiency for data providers and users in enabling more flexibility in the provider pipelines; long term efficiency improvement in the archive with standard formats |



# Design Discions (2

| | Resources | Training | Usability | Efficiency |
|---|---|---|---|---|
| Distributed Infrastructure | Major short-term impact will be on system development; Improved efficiency, tracking, access to data realized | Primary training issue is in developing software which can access online services over the network | Better support for design, validation and submission to PDS; more unified PDS-wide access to data; International access and support will be enabled. | Improved efficiency by reducing the number of versions of ad hoc tools and information |
| Format Transformation Tools and Services | Primary impact will be on developing tools to convert to/from PDS. | Primary training issue is in developing software that can perform transformation to/from PDS4 | Improved usability by allowing users to access data in user preferred formats. | Improved efficiency for data providers and users in enabling more flexibility in the provider pipelines; long term efficiency improvement in the archive with standard formats |



# Design Decisions (3)

| | Resources | Training | Usability | Efficiency |
|---|---|---|---|---|
| Distributed Infrastructure/ Service Oriented Architecture | Major short-term impact will be on system development; Improved efficiency, tracking, access to data realized | Primary training issue is in developing software which can access online services over the network | Better support for design, validation and submission to PDS; more unified PDS-wide access to data; International access and support will be enabled. | Improved efficiency by reducing the number of versions of ad hoc tools and information |
| Format Transformation Tools and Services | Primary impact will be on developing tools to convert to/from PDS. | Primary training issue is in developing software that can perform transformation to/from PDS4 | Improved usability by allowing users to access data in user preferred formats. | Improved efficiency for data providers and users in enabling more flexibility in the provider pipelines; long term efficiency improvement in the archive with standard formats |

# Design Recommendations under preparation

- Adopt simple structures at the data object level for insertion into the archive, however,
  - Improve data provider delivery efficiency by supporting translation into primitive archive formats
  - Support transformation to popular formats on distribution
  - Ensure simple formats are easy to convert and preserve long term
- Adopt ISO/IEC 11179 as a mechanism to structure and improve the data dictionary
  - Add namespae to partition the data dictionary
  - Limit keyword changes at the globel model; keyword changes are moved to different name spaces for mission, node, etc



# More design decisions

- Move away from ODL
  - PDS is aware of the problems people are encountering; it requires fixing
  - There, PDS evaluated PDS vs XML
  - We did a pilot experience to demonstrate how a PDS4 XML document can be easily converted to a PVL/ODL view for display
  - XML seems to be ther ight decisoin and folks are warming up to it; more on the liater.
- Adopt a service oriented architecture
  - It will reduce cost
  - It will improve interoperability
  - It will enable PDS to better work with IPDA
- Format transformation
  - Enable toPDS functions and fromPDS functions to accept and deliver data in certain formats that might be optimized for the domain (mission, science analysis, etc)
  - What we can do here is a function of our budget, but will technically scale



```
OBJECT = PRODUCT_IMAGE_GRAYSCALE;

  OBJECT = IDENTIFICATION_SECTION;
    DD_VERSION_ID          = "DD_VERSION_ID";
    IDENTIFIER             = "PDS4_IMG_IMAGE_GRAYSCALE_ID:V1.0";
    LABEL_REVISION_NOTE    = "20090101:1.0 - initial version;
                              20090102:1.1 - added another column";
    PDS_VERSION_ID         = "PDS4.0";
    PRODUCT_CREATION_TIME = 1998-07-14T00:36:08.000;
    TITLE                  = "MARS PATHFINDER LANDER Experiment";
    URN                    = "http://URN:MPFL-M-IMP-2-EDR-V1.0:PDS4_IMG_GRAY-SCALE_IMAGE_ODL.LBL:1.0";
    VERSION                = "1.0";
  END_OBJECT = IDENTIFICATION_SECTION;


  OBJECT = DESCRIPTION_SECTION;
    DESCRIPTION            = "ANNOTATION FOR THE PRODUCT GOES HERE.";
  END_OBJECT = DESCRIPTION_SECTION;


  OBJECT = CIRCUMSTANCES_OF_OBSERVATION_SECTION;
    COMMENT                = "Observation Intent";
    SPACECRAFT_CLOCK_START_COUNT = "1246943630";
    SPACECRAFT_CLOCK_STOP_COUNT  = "N/A";
    START_TIME                   = "N/A";
    STOP_TIME                    = "N/A";
  END_OBJECT = CIRCUMSTANCES_OF_OBSERVATION_SECTION;

  OBJECT = DATASET_SECTION;
    DATA_SET_ID            = "MPFL-M-IMP-2-EDR-V1.0";
  END_OBJECT = DATASET_SECTION;


  OBJECT = MISSION_SECTION;
    MISSION_NAME           = "MARS PATHFINDER";
  END_OBJECT = MISSION_SECTION;


  OBJECT = INSTRUMENT_HOST_SECTION;
    INSTRUMENT_HOST_ID     = "MPFL";
  END_OBJECT = INSTRUMENT_HOST_SECTION;


  OBJECT = INSTRUMENT_SECTION;
    INSTRUMENT_ID_NEW      = "IMP";
  END_OBJECT = INSTRUMENT_SECTION;


  OBJECT = NODE_SECTION;
    NODE_NAME              = "IMAGING";
  END_OBJECT = NODE_SECTION;

OBJECT = TARGET_SECTION;
    TARGET_NAME            = "MARS";
  END_OBJECT = TARGET_SECTION;


  OBJECT = TAGGED_IMAGE_GRAYSCALE_SET;

    OBJECT = IMAGE_GRAYSCALE;
      DATA_LOCATION        = ("MPFL_M_IMP_IMAGE_FILE",1);

      LOCAL_IDENTIFIER     = "MPFL_M_IMP_IMAGE";
```



```
– <product_label>
    <pds_version_id>PDS4.0</pds_version_id>
    <dd_version_id>DD_VERSION_ID</dd_version_id>
    <identifier>PDS4_IMG_IMAGE_GRAYSCALE_ID:V1.0</identifier>
    <title>MARS PATHFINDER LANDER Experiment</title>
  – <urn>
      http://URN:MPFL-M-IMP-2-EDR-V1.0:PDS4_IMG_GRAY-SCALE_IMAGE_ODL.LBL:1.0
    </urn>
    <version>1.0</version>
  – <label_revision_note>
      20090101:1.0 - initial version; 20090102:1.1 - added another column
    </label_revision_note>
    <product_creation_time>1998-07-14T00:36:08.000</product_creation_time>
    <description>ANNOTATION FOR THE PRODUCT GOES HERE.</description>
    <comment>Observation Intent</comment>
    <spacecraft_clock_start_count>1246943630</spacecraft_clock_start_count>
    <spacecraft_clock_stop_count>N/A</spacecraft_clock_stop_count>
    <start_time>N/A</start_time>
    <stop_time>N/A</stop_time>
    <data_set_id>MPFL-M-IMP-2-EDR-V1.0</data_set_id>
    <mission_name>MARS PATHFINDER</mission_name>
    <instrument_host_id>MPFL</instrument_host_id>
    <instrument_id_new>IMP</instrument_id_new>
    <node_name>IMAGING</node_name>
    <target_name>MARS</target_name>
  – <array>
      <data_location byte_offset="1">MPFL_M_IMP_IMAGE_FILE</data_location>
      <local_identifier>MPFL_M_IMP_IMAGE</local_identifier>
      <axes_order>FAST2SLOW</axes_order>
      <byte_order>MSBF</byte_order>
      <file_type>BINARY</file_type>
      <first_element>TOPLEFT</first_element>
      <min_index>0</min_index>
      <number_of_axes>2</number_of_axes>
      <element_bytes>2</element_bytes>
      <element_offset>N/A</element_offset>
      <element_scaling_factor>N/A</element_scaling_factor>
```

# Migration Options

- No Migration
  - Users continue to have access to PDS3 and PDS4 data
  - Software/tools are in place to support both PDS3 and PDS4
- On-demand Migration
  - Users continue to have access to PDS3 and PDS4 data
  - Software/tools are in place to support both PDS3 and PDS4
  - Some data sets migrated and redelivered to NSSDC
- Full Migration
  - Users have access to data in PDS4 data format only
  - Users Software/Tools only support PDS4
  - All data migrated and redelivered to NSSDC
  - Accept PDS3 and PDS4 data from data providers



# Migration Options

| | Resources | Training | Usability | Efficiency |
|---|---|---|---|---|
| No Migration | No impact on resources to convert data, however, PDS software will need to support PDS3 and PDS4 | Users will need to be capable of working with PDS3 and PDS4 data | Limited support for working with PDS3 data in the future | Most cost-effective solution |
| On Demand Migration | Impact on conversion of a subset of critical data sets; PDS software will need to support PDS3 and PDS4 | Users will need to be capable of working with PDS3 and PDS4 data | Limited support for working with PDS3 data in the future; critical data sets will be converted to PDS4 to improve usability | More costly than "No Migration", however, improves usability following a pragmatic approach |
| Full Migration | Substantial impact in converting data, redelivering to NSSDC, and developing supporting software | Users will ultimately need to only be familiar with PDS4 | Usability would be improved since PDS data will be brought up to date | Substantial costs in migrating all data |



# Stakeholder Impacts

| | Data Providers | Data Users | Discipline Nodes | Engineering Node |
|---|---|---|---|---|
| No Migration | No impact | Users will need to be familiar with PDS3 and PDS4 data formats | Nodes will need to continue to provide support for PDS3 data | Engineering Node must continue to provide software support for PDS3 and PDS4 |
| On Demand Migration | No impact | Users will need to be familiar with PDS3 and PDS4 data formats. However, critical data sets can be migrated to enhance usability. | Nodes will need to continue to provide limited support for PDS3 data; minor impact in migrating critical data sets | Engineering Node must continue to provide software support for PDS3 and PDS4 |
| Full Migration | No impact | Data users will need to eventually only learn PDS4 | Substantial impact in migrating and redelivering all data to PDS4 | PDS3 tools and services can be retired once migration is complete |



# Transition Options

- Big bang
  - Entire PDS 2010 system is stood up and then PDS transitions
- Incremental
  - Minimal PDS 2010 system is put in place and then PDS transitions
  - Increasing levels of capability are put in place over time to support enhanced user services and greater efficiency
  - Acceptance of PDS4 data can occur in stages (e.g., certain missions, etc)
- Both can address continued support for PDS3

# Transition Matrix

| | Resources | Training | Usability | Efficiency |
|---|---|---|---|---|
| Big Bang | Resources will be fixed, however, the schedule can scale which can delay the release of PDS 2010 | Not a major transition issues | If all services are in place first, then this will increase usability. However, deployment of PDS 2010 could be delayed longer affecting the ability to put PDS4 model in place | Less efficient of the two approaches since it delays rolling out PDS 2010 |
| Incremental | Resources and the timeline can be scaled with PDS budget (whether overguide is provided or not) | Not a major transition issues | Critical capabilities are put in place earlier, but the trade off is number of functional capabilities that must be supported prior to ingestion/distribution | More efficiency of the two approaches since it allows for delivery of increasing capability and the MC can then determine functional capabilities required in order to accept and distribute PDS4 data |

# Transition Impacts

| | Data Providers | Data Users | Discipline Nodes | Engineering Node |
|---|---|---|---|---|
| Big Bang | Capabilities and services are delayed, however, greater functionality would be in place when they begin to deliver PDS4 data | Data users would get data in PDS4 format later, but could potentially have more tool support | Larger impact on the nodes if all nodes need to put services and capabilities in place first | Larger impact on the Engineering Node to put services and tools in place and ensure PDS-wide readiness to accept data for all new missions in PDS4 at once |
| Incremental | Capabilities can be put in place earlier and begin flow of PDS4 data and support | Data users would get data in PDS4 format earlier, but would possibly have less tool support | Less impact on the nodes if data is accepted in stages | Less impact on the Engineering Node and improved project performance since results can be realized earlier and PDS can deploy services and train over time |



# Build Plan…Incremental

| | Description | Impact | Timeline |
|---|---|---|---|
| **Build 2010-1** (Draft Capabilities) | Initial infrastructure services released; draft release of PDS Standards Reference | Principal impact is on Engineering Node for the infrastructure release and the DDWG to release draft standards reference | |
| **Build 2010-2** (Minimal Infrastructure to support PDS4 data ingestion and distribution) | PDS4 Data/Information Model release and standards reference<br>Data Dictionary Service<br>Upgraded validation tool<br>Standard grammar<br>Catalog registry<br>Search/web portal changes | Principal impact will be getting tools and service and standard reference in place; PDS MC will need to determine phasing plan for acceptance of data | |
| **Build 2010-3** **(**Increased User Capabilities) | to/from PDS transformation libraries and service<br>Reconfigured web infrastructure to integrate with product registry<br>Improved search across PDS | Improved services will be deployed to nodes including tracking and access to data products | |

# IPDA Opportunities

- PDS wants IPDA involvement, particularly data standards
- PDS is wants to share
- IPDA conceptual architecture is already in alignment
  - Specifics will be in the standards

# Questions?



# Backup

# Guiding Information for Design

- Roadmap
  - http://pds-engineering.jpl.nasa.gov/projects/PDS4/Exchange/PDS_Roadmap.pdf
- PDS Level 1, 2, 3 (System-Level)
  - http://pds-engineering.jpl.nasa.gov/index.cfm?pid=5&cid=72
  - Really, not a PDS3 set of requirements
- PDS4 Concept Papers
  - http://pds-engineering.jpl.nasa.gov/index.cfm?pid=100&cid=119 (Architecture)
  - http://pds-engineering.jpl.nasa.gov/index.cfm?pid=100&cid=120 (Data Model)
  - http://pds-engineering.jpl.nasa.gov/index.cfm?pid=100&cid=121 (User Support)
- PDS Vision and Exec Summary
  - http://pds-engineering.jpl.nasa.gov/projects/PDS4/pds2010-execsummary20080701.pdf
- PDS 2010 Architecture
  - http://pds-engineering.jpl.nasa.gov/index.cfm?pid=100&cid=131 (System Architecture)
  - http://pds-engineering.jpl.nasa.gov/index.cfm?pid=5&cid=125 (Data Architecture)