PDS4 and Build 5a Update

Dan Crichton, Emily Law
November 2014
PDS4 and Related MC Topics

- PDS4 Report and Build 5a – Dan Crichton and Emily Law
- IM/DDWG – Steve Hughes
- Software – Sean Hardman
- Tool Planning – Dan Crichton
- AMMOS Pipeline Service – Costin
- ARC Tool Updates – Mark Rose
- Discussion
PDS4: The Next Generation PDS

• 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 validation and discovery
  • 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 open source registry infrastructure to track and manage every product in PDS
  • A distributed search infrastructure
Challenge: End-to-End System and Data Integration

Core PDS4

Data Providers → Transform → Ingest → PDS Data Management → Distribution → Transform → Users

- 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
**Information Architecture Concepts**

**Design/change starts here**

---

**Product**

**Tagged Data Object** (Information Object)

```xml
<Array_2D_Image>
  <local_identifier>MPFL-M-IMP_IMG_GRAYSCALE…</local_identifier>
  <offset unit="byte">0</offset>
  <axes>2</axes>
  <axis_index_order>Last Index Fastest…
  <Element_Array>
    <data_type>UnsignedMSB2</data_type>
    <unit>data number</unit>
  </Element_Array>
  <Axis_Array>
    <axis_name>Line</axis_name>
    <elements>248</elements>
    <unit>not applicable</unit>
    <sequence_number>1</sequence_number>
  </Axis_Array>
  <Axis_Array>
    <axis_name>Sample</axis_name>
    <elements>256</elements>
    <unit>not applicable</unit>
    <sequence_number>2</sequence_number>
  </Axis_Array>
</Array_2D_Image>
```

---

**Describes**

**Data Object**
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
Summary of Progress to Date

• Phased-approach to building PDS4; learning over time.
• Initial requirements in place
• PDS-wide system architecture defined
• Major reviews conducted (Design Review 1 and 2, ORR)
• System builds grouped by purpose: build 1, 2, 3 and 4
  • Iteratively increase capability and stability
  • Allows DNs and missions to lock into PDS4 versions when capabilities exist
• Operational capabilities deployed
  • EN fully running PDS4 software supporting access to both PDS3 and PDS4 services at nodes mitigating migration pressure
• Change control board established
  • JIRA deployed to manage tracking
• Product development underway at nodes and internationally
  • Initial peer reviews conducted
• IPDA endorsement of PDS4
• LADEE now in the middle of using PDS4
# Project Lifecycle thru Build 3

<table>
<thead>
<tr>
<th>Project Lifecycle</th>
<th>Pre-Formulation</th>
<th>Formulation</th>
<th>Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gates &amp; Major Events</td>
<td>KDP: Study</td>
<td>KDP: Project Plan &amp; Arch</td>
<td>KDP: Prelim Design</td>
</tr>
<tr>
<td></td>
<td>Begin Study Project</td>
<td>Project Plan PDS4 Prelim Architecture</td>
<td>Build 1: Prototype build</td>
</tr>
<tr>
<td></td>
<td>Study/Concepts</td>
<td>PDS4 Design</td>
<td>1a (Oct 2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1b</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1c (Aug 2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1d</td>
</tr>
</tbody>
</table>

| | PDS External System Design Review I (Mar 2010) | Build 1b PDS Stds Assessment (Dec 2010) |
| | PDS External System Design Review II (June 2011) | Build 1c IPDA Stds Assessment (Apr 2011) |
| | ORR (Start Label Design) (LADEE/MAVEN) (Nov 2011) | Build 1d External Stds Assessment (Aug 2011) |

Architecture, requirements, design, test, releases posted at: [http://pds-engineering.jpl.nasa.gov](http://pds-engineering.jpl.nasa.gov)
Build 4 and 5

### Implementation

- **Build 3**: V1.0 of PDS4 Standards; Transition EN to PDS4 Software
  - 3a (Sep 2012)
  - 3b (Mar 2013)

- **PDS MC Review of Build 3b** (April 2013)

- **KDP: Release V1.0 of PDS4 Data Standards**
  - 4a (Sep 2013)

- **KDP: Deploy for LADEE/MAVEN**
  - Build 4: Release to Data Users (LADEE/MAVEN); Deploy PDS4 Software to DNs
  - 4b (Mar 2014)
  - 5a (Sep 2014)
  - 5b (Mar 2015)

- **Phoenix Beta Test** (Dec 2012)

- **Operational Readiness Review (LADEE/MAVEN Deployment)** (September 2013)

- **Build 5**: Expand to future mission sets at NASA, ESA, JAXA, etc

---

*PDS4 now under CCB process*
The System Architecture presentation will map these to LADEE and MAVEN.
The Information Model Driven Process & Artifacts

Requirements & Domain Knowledge

Protégé Ontology Modeling Tool

PDS4 Information Model

Filter and Translator

XML Schema (pds)

PDS4 Data Dictionary (ISO/IEC 11179)

XML Document (Label Template)

PDS4 Data Dictionary (Doc and DB)

Information Model Specification

Query Models

Registry Configuration Parameters

XMI/UML

This is given to data providers (e.g., LADEE/MAVEN)
Build 5a

• Following the lifecycle we have established, we are in process of the Build 5a release
  • Information Model and Software delivered to I&T on Sep 30, 2014
  • Completed I&T (that integrate tools, services and data products including regression test) on Oct 27
  • Executed 41 test cases, no major anomaly

• Deployment at EN has began
Build 5a Software Components

- Information Model V1.3.0.1 that includes SCRs approved by the CCB
- Software
  - Generate, Transform, Validate
  - Harvest, Registry
  - Catalog Ingestion
  - Search (Core, Service, UI, Product Search UI, Data Set View)
  - Storage
  - Security
  - Transport
  - Report
Build 5a Documents and Schema

  - Schemas
- [https://pds-engineering.jpl.nasa.gov/development/pds4/5.0.0/release/index-5.0.0.html](https://pds-engineering.jpl.nasa.gov/development/pds4/5.0.0/release/index-5.0.0.html)
  - Software Release Description, Installation Procedures
- [https://pds-engineering.jpl.nasa.gov/content/build-5a-deliverables](https://pds-engineering.jpl.nasa.gov/content/build-5a-deliverables)
  - I&T Test Reports, Test Data, Requirements Traceability
PDS4 Planned Mission Support

LADEE (NASA)  InSight (NASA)  BepiColumbo (ESA/JAXA)

MAVEN (NASA)  Osiris-REx (NASA)  ExoMars (ESA/Russia)  JUICE (ESA)

...also Hyabussa-2, Chandryaan-2

Endorsed by the International Planetary Data Alliance in July 2012 –
https://planetarydata.org/documents/steering-committee/ipda-endorsements-recommendations-and-actions
Legacy PDS3 Implementation

Current Missions

- PDS3 Pipeline
  - Datasets + Products
- PDS3 Ingest
- PDS3 Central Catalog
- PDS3 Archive @ DNs
- PDS3 Services
  - PDS3 Metadata Index
- PDS Portal

Users
On-going Support for PD3 Missions (e.g., Cassini, Dawn, MSL, etc)

1. PD3 infrastructure deployed at EN; Central catalog migrated.
Support for LADEE/MAVEN/…

Current Missions

PDS3 Pipeline

Datasets + Products

PDS3 Ingest

PDS3 Central Catalog

PDS3 Archive @ DNs

PDS3 Services

PDS3 Data Migration (label, label+data)

(1) *PDS4 infrastructure deployed at EN; Central catalog migrated.*
(2) *Minimized migration pressure.*
(3) *Working towards acceptance/distribution of new PDS4 mission data*

Updated PDS Portal

PDS4 Ingest

PDS4 Archive @ DNs

PDS4 Registry

PDS4 Archive

PDS4 Services

PDS4 XML Label

PDS4 Metadata Index

toPDS4 Transform

Users

NOTE: PDS3 Services phased out overtime

PDS4 Pipeline

Build 4

(PDS3, MAVEN, O-Rex, Insight)

New Missions
International Support
IPDA

- PDS involved in many projects
- PDS4 is the primary collaboration vehicle
- New collaboration projects kicked off
  - Identified at Moscow meeting
  - Major focus are other agency implementations of PDS4 and their assessments
- October telecon included discussions on
  - Cartography
  - Geometry

http://planetarydata.org
PDS4 Policies

• The following PDS4 specific policies are posted to http://pds.nasa.gov/policy
  • PDS Policy on Formats for PDS4 Data and Documentation (June 2014)
  • PDS Policy on Data Processing Levels (March 2013)
  • PDS Policy on Transition from PDS3 to PDS4 (November 2010)
Updated Validation Definition

Procedure for Validating PDS4 Data Deliveries

Prior to delivery to the PDS, data providers shall validate PDS4 bundles, collections, and/or basic products using PDS supplied XML Schema and Schematron files and validation tools to ensure syntactic, semantic, referential integrity, and specific content validity. Data providers should also use visual inspection to validate content that cannot be done programmatically (i.e., by using product content visualization tools).

The following validation should be performed before delivery:

Product Label

1. Syntactic validation: verify the XML label is syntactically valid.
2. Semantic validation: verify the XML label is semantically valid.
3. Referential integrity:
   a. verify that the relationships described, in and between digital objects described in the XML label, are consistent and represented.
   b. verify that the relationships described between products are consistent and represented.
4. Content validation: ensure the XML label accurately describes the data product.

Other Files

1. Mission Dictionaries: Ensure that the XML Schema and Schematron files provided by the mission are valid.

• Discussed at last MC
• Crichton took action to work with Simpson
Where we need help

• Registry population from PDS3 archives

• Report service population

• Identifying tool priorities (this will come later)
  • Included is a proposal of how to continue to inject node input in the development process as PDS4 matures

• Supporting tool and software sharing
  • Registration and curation of available tools
Backup
Schedule

• [https://pds-engineering.jpl.nasa.gov/sites/default/files/pds4-project-schedule.pdf](https://pds-engineering.jpl.nasa.gov/sites/default/files/pds4-project-schedule.pdf)
Approach

• Testing
  • Perform against level 5 requirements defined in Software Requirements and Design Documents (SRD/SDD)
  • Test software under CM using Subversion
  • Perform on test environment
  • Perform regression test over time and on incremental builds

• Documentation
  • Release descriptions
  • Test procedure and report
  • Update for increment builds
Process

• Software Test per system build
  • Unit Test perform by EN development staff
  • Integration Test perform by EN operations staff following Test Plan and Procedures

• Defects and issues
  • Utilize JIRA for issue tracking

• Report
  • Collect metrics (e.g., # of defects, test cases)
  • Document test results