J. Mafi

PDS Tech Session

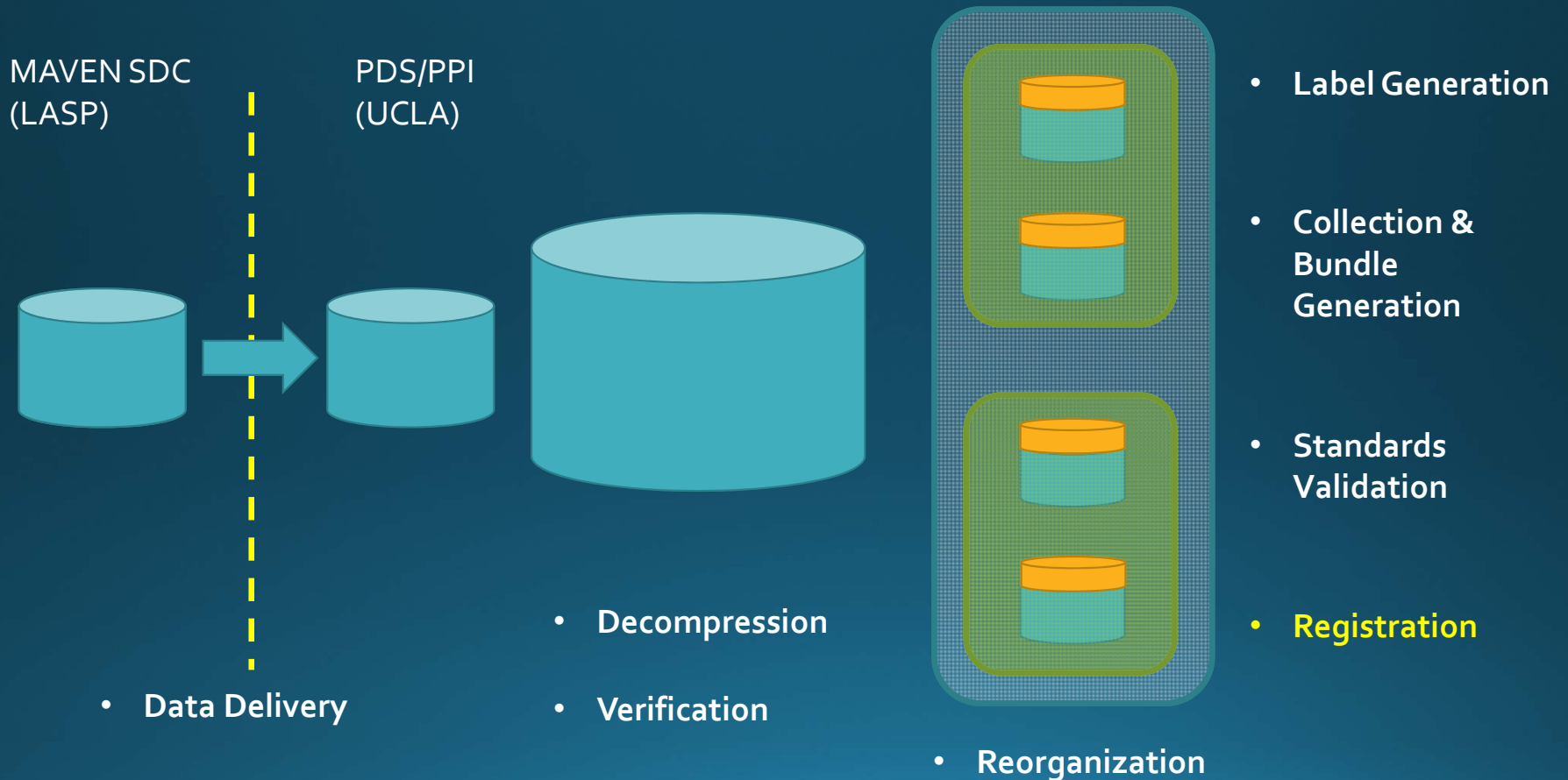21 September 2016

MAVEN Archiving Notes

Overview

- MAVEN Archive
- PPI Archive Procedure
  - Processes
  - Tools
- Tool Needs

MAVEN Archive Summary

- Atmospheres Holdings
  - 3 instruments
    - 6 data bundles (with additional document and calibration bundles)
    - 30 data collections (25 FITS, 5 ASCII Delimited)
  - Instrument teams produce PDS4 metadata

- PPI Holdings
  - 7 instruments; 3 non-instrument data providers
    - 12 data bundles (with additional document bundles)
    - 85 data collections (64 CDF, 19 ASCII Table, 2 ASCII Delimited)
  - PPI produces PDS4 metadata for 11 bundles (82 collections)
  - 1 additional bundle in design: raw telemetry (Product_Native); will be safed
  - Possible Radio Science data archive has been discussed

Procedure Details

Process	Description	Tools
Data Delivery	• Data collected at the MAVEN Science Data Center (SDC) • Delivered as tar/gzip to PPI once cleared by the instrument teams	sftp
Decompression	• Data decompressed into the directory structure used at the MAVEN SDC	tar (GNU)
Verification	• MD5 checksums verified	md5sum/ igpp.md5check
Reorganization	• Data reorganized into the PDS archive structure	Perl script
Label Generation	• PDS4 labels are generated from CDF metadata, pseudo PDS3 labels, and lookup tables	pds.cdf, igpp.docgen
Collection & Bundle Generation	• Collection inventory files (CSV), and labels generated • Readme.txt and bundle label files generated	Perl scripts, manual processes
Standards Validation	• Archive XML files checked for PDS4 standards compliance	oXygen, Validate Tool + ParseValidate
Registration	• Archive registered in the PDS Registry	Registry Tool

Tool Needs

Collection & Bundle Generation	• Collection inventory files (CSV), and labels generated • Readme.txt and bundle label files generated	Perl scripts, manual processes
Standards Validation	• Archive XML files checked for PDS4 standards compliance	oXygen, Validate Tool + ParseValidate

1. Collection label generation tool

2. Data format validation tool

3. Logical Identifier validation tool

Tool Needs: Collection Label Tool

Process	Description	Tools
Collection & Bundle Generation	• Collection inventory files (CSV), and labels generated • Readme.txt and bundle label files generated	perl scripts, manual process

- Issue:
  - Collection label generation for accumulating data sets involves a manual effort for a process that could be automated.

- Need:
  - A tool to interact with Registry to collect individual data product metadata for inclusion in the collection metadata
  - Designed to work with label generation software

- Priority:
  - Medium

Tool Needs: Data Format Validation

Process	Description	Tools
Standards Validation	• Archive XML files checked for PDS4 standards compliance	oXygen, Validate Tool

- Issue:
  - Validation tools do not check accuracy of data structure description in the label

- Need:
  - Verify that the data structure described in the PDS4 label files matches the structure of the data file

- Priority:
  - High

Tool Needs: Logical Identifier Validation

Process	Description	Tools
Standards Validation	• Archive XML files checked for PDS4 standards compliance	oXygen, Validate Tool

- Issue:
  - Validation tools do not check product and referenced LIDs against the Registry
- Need:
  - Verify that LIDs/LIDVIDs for new products do not already exist
  - Verify that LIDs/LIDVIDs referenced in labels do exist
  - Verify that LIDs reference in bundle and collection inventories exist
- Priority:
  - Medium/High

Backup Slides

MAVEN Archive Overview

Instrument	Data Bundles (Collections)	Format(s)	Label Prod.	Notes
EUV	2 (1/2)	CDF	PPI	
LPW	3 (17/8/2)	CDF	PPI	
MAG	1 (3)	ASCII Table	MAG	
SEP	1 (3)	CDF	PPI	
STATIC	1 (20)	CDF	PPI	
SWEA	1 (5)	CDF	PPI	
SWIA	1 (6)	CDF	PPI	
Ancillary	1 (17)	ASCII Table; CSV	PPI	
KP	1 (1)	CSV	PPI	
Raw TLM	???	Binary Table	PPI	In design

Tool Descriptions

Step	Tool	Description	Developer
Verification	igpp.md5check	Calculates MD5 checksums; compares checksums to those in a log file	UCLA
Label Generation	igpp.docgen	Populates an Apache Velocity template using metadata scanned from an PDS label, XML, or CDF file producing a PDS label, XML, or plain text file	UCLA
	pds.cdf.CDF	Reads and returns CDF metadata	UCLA
Standards Validation	oXygen, etc.	XML editor and debugger	COTS
	Validate Tool	Validates PDS4 product labels and product data	JPL
	ParseValidate	Perl script to reduce Validate Tool report	UCLA
Registration	Registry Service	Software for tracking, auditing, locating, and maintaining artifacts within the system	JPL