

Introduction to PDS4

MOSES MILAZZO

EUROPA CLIPPER PSG MEETING

PASADENA, CA

JUNE 15, 2018

PDS4 Training Schedule for ECM

Europa Clipper PSG PDS4 Training Session Agenda

Friday June 15, 2018

Time (PDT)	Duration (min)	Topic	Speaker
8:00 AM	60	Introduction to PDS4 (40 min presentation, 20 min questions/discussion)	Moses Milazzo (USGS - PDS IMG Node)
9:00 AM	30	Discipline and Mission Dictionaries (20 min presentation, 10 min questions/discussion)	Ed Guinness (Washington University - PDS GEO Node)
9:30 AM	30	PDS4 Tools Overview (20 min presentation, 10 min questions/discussion)	Michael Cayanan (JPL - PDS Engineering Node)
10:00 AM	15	Break	
10:15 AM	30	Using PLAID for PDS4 Label Design (20 min presentation/demo, 10 min questions/discussion)	Stirling Algermissen (JPL - MGSS)
10:45 AM	30	APPS for PDS4 Bundle/Collection Development (20 min presentation, 10 min questions/discussion)	Kate Crombie (Indigo Information Services)
11:15 AM	30	Logistics when working with PDS mission POCs (20 min presentation, 10 min questions/discussion)	Moses Milazzo (USGS - PDS IMG Node)

WebEx Information

<https://jpl.webex.com/jpl/j.php?MTID=me5c5f3e8dd511e470a83a102ac5342f7>

Meeting number: 900 205 617

Overview

Review PDS4 basics

What is PDS4?

Components of a PDS4 archive

Structure of PDS4 metadata

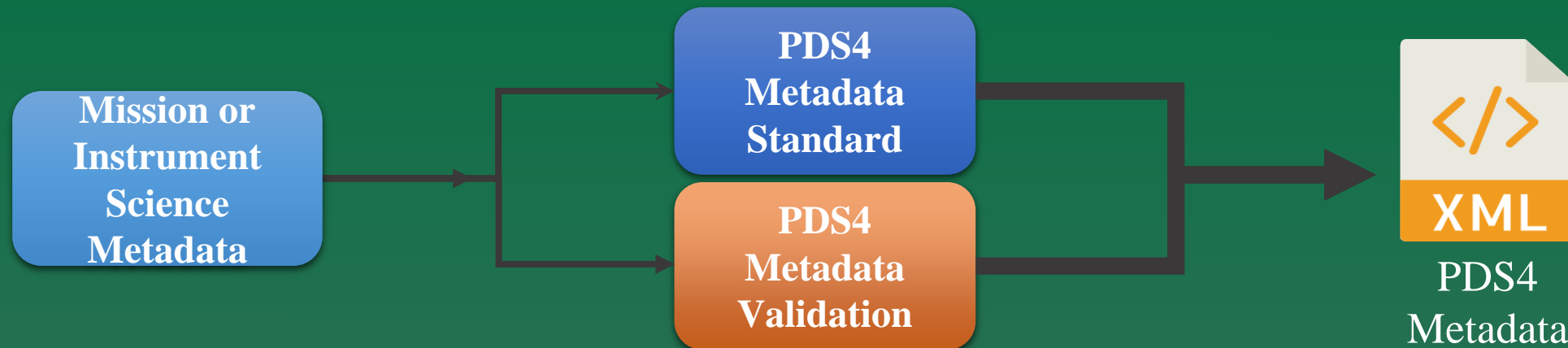
Designing an Archive, Editing a Label

Validation, Tools, Data Dictionaries

Logistics of Clipper PDS4 Development

<https://pds.nasa.gov/>

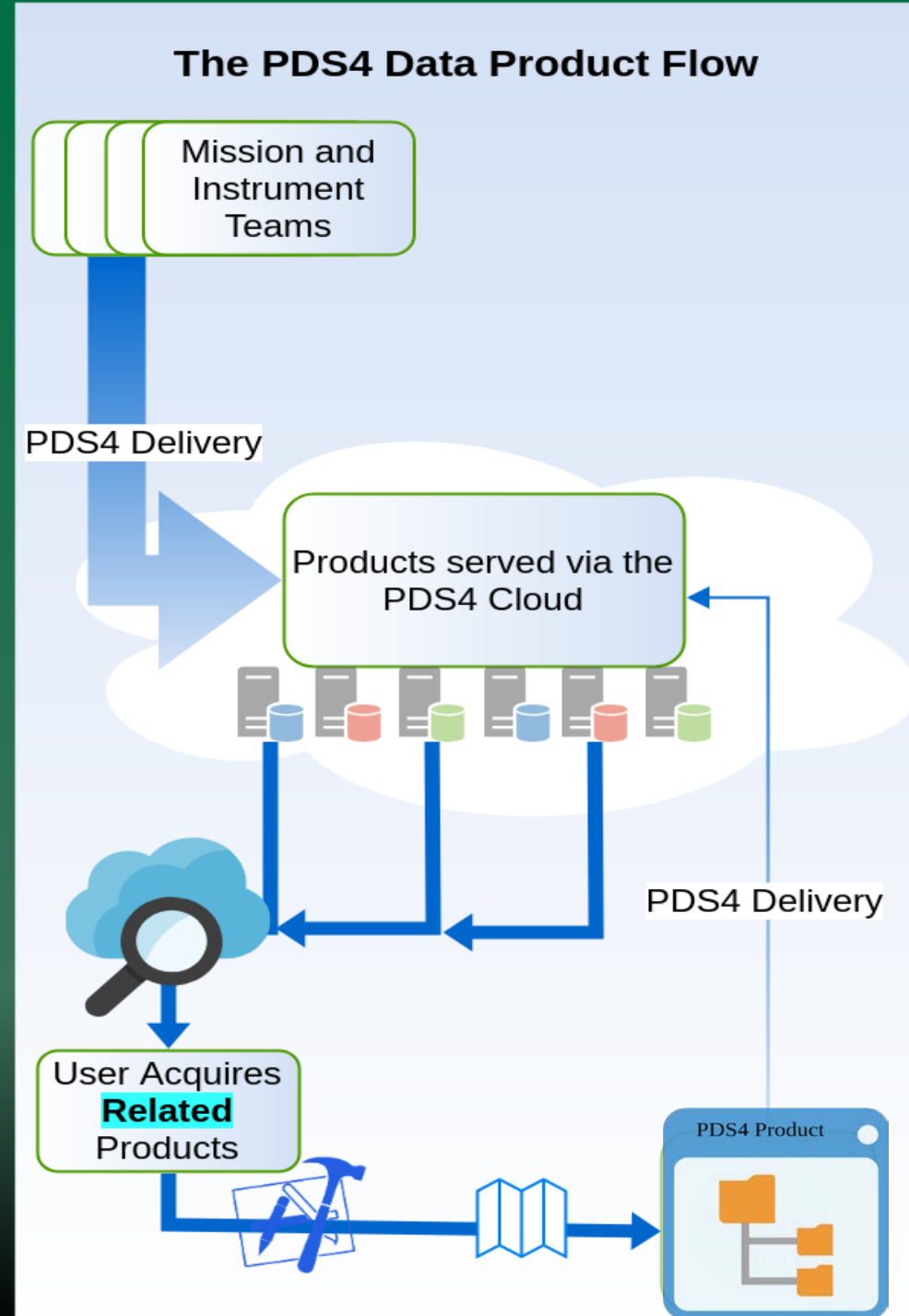
What is PDS4?



- *PDS4* is a modernization from PDS version 3 to 4
- Based on modern information modeling
 - New archival standards
 - New database infrastructure
- NASA requirement since 2011 for new deliveries

What is PDS4?

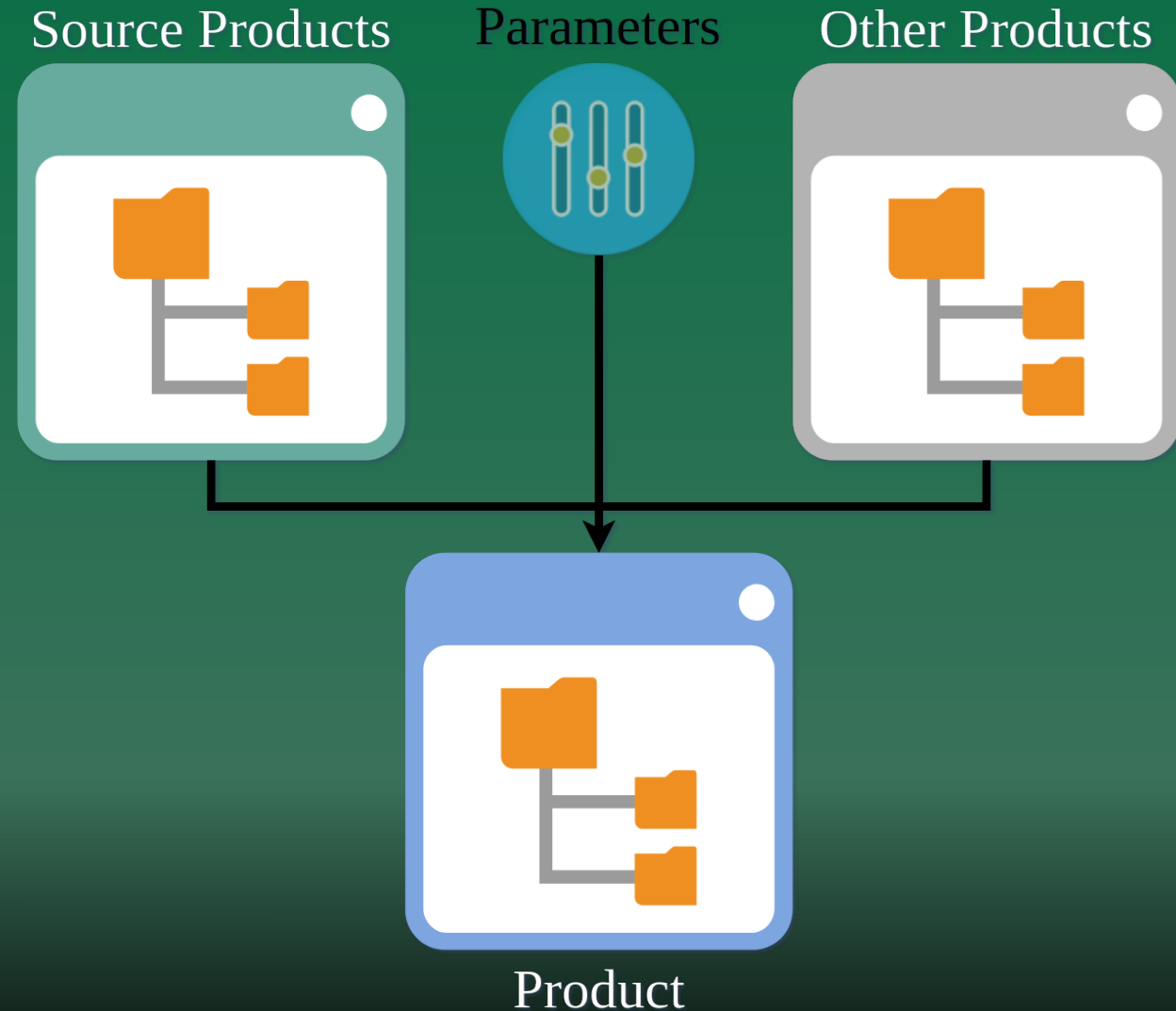
- Improves efficiency of
 - Data ingestion
 - Tracking
 - **User** access to data:
 - Search
 - Distribution



What is PDS4?

- PDS4 is designed to improve **user access to data**
 - Supports explicit referencing related material
 - May link to generating parameters
 - May reference other products
 - Internal or external references are optional

The power of references is only realized if they are included!



PDS4 Information Model

- PDS4 information model:
 - All PDS4 data fit within a common model to improve:
 - Validation
 - Discovery
 - Usage
- An explicit software/technical architecture
 - Distributed services: PDS nodes, subnodes, international partners
 - Consistent protocols improve access to data, services
 - Registry allows tracking, managing all products in PDS

PDS4 Information Model

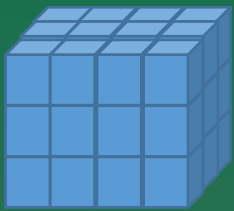
- Defines explicit relationships within PDS
- PDS4 metadata for the core dictionary is well-defined
 - Extended to discipline and mission level dictionaries
- Consistency in PDS labels across many instruments and observation types
- Provides a single authoritative source for data standards
- Based on international standards for data dictionaries

PDS4: XML in two parts

- Schema defines:
 - Attributes, order and content
 - Data types
 - Structure of classes and attributes
 - Required and optional classes and attributes
- Schematron:
 - Provides additional rules
 - Used for standard value lists
 - Enforces context-dependent constraints (“If this, then that”; Either this or that, but not both”)
- The PDS4 core schema is derived from the information model
 - <https://pds.jpl.nasa.gov/datastandards/schema/>

PDS4: Fundamental Data Structures

PDS4 supports science and support data describable using one of these fundamental data structures:



Array – Homogeneous array of scalars: Images, cubes; up to 16 dimensions (all elements must have the same data type)

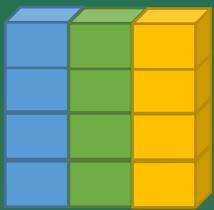
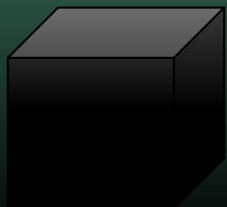


Table – ASCII or binary data with a repeating structure, fixed-width fields



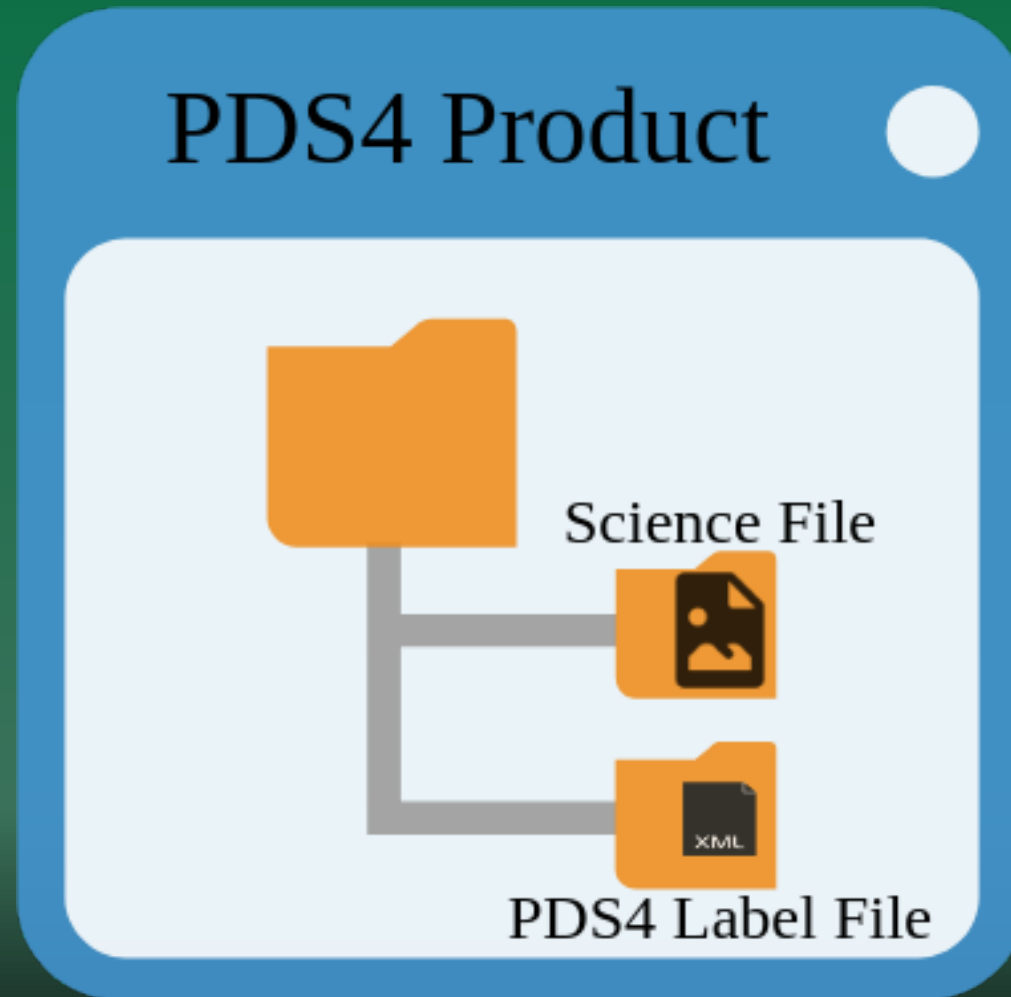
Parsable Byte Stream – Bytes formatted with standard parsing rules: ASCII data with a repeating record structure, variable width fields with a delimiter (e.g. CSV table, text file, XML file)



Encoded Byte Stream – Uses certain well-known standards (e.g. PDF/A, JPEG). **NOT USED FOR SCIENCE FILES!**

PDS4: Products

- **PDS Label:**
 - xml file with PDS4 metadata
 - validated through PDS4 Schema and Schematron
- **PDS Product:**
 - A PDS label + files it describes
 - PDS4 labels are co-located with the files that they describe
- A PDS4 Product is **not**:
 - A lonely PDS4 label*
 - A lonely science or other file



PDS4: Product Types



- **Basic Product:** Science file(s) and associated PDS4 label



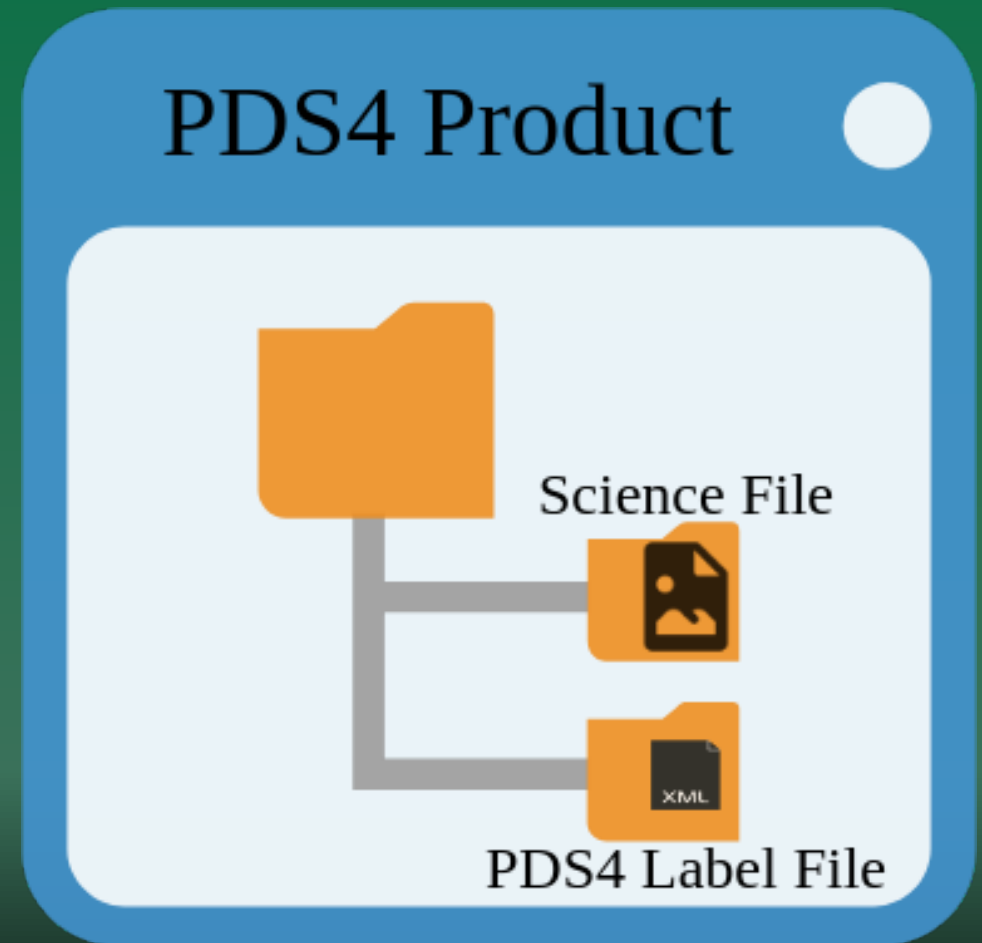
- **Product Collection:** Related Basic Products of the same type



- **Product Bundle:** Related Collections

Basic Products

- One or more science files
- A PDS label file describing the science file(s)
- Many types (data, tables, browse, documents, etc.)
- Basic products are frequently referred to informally as “products”



Collection Products

- **Inventory.csv**
 - A table file that lists all Basic Products that are part of this Collection

- **PDS4 Label.xml**
 - Describes Inventory.csv
 - The collection label file may optionally roll-up any metadata contained in the individual member products



Bundle Products

- **PDS4 Label:** Describes the Collections in the Bundle
- Optional Readme .txt
 - Must either be ASCII text or UTF-8
- Overview of bundle content and organization.

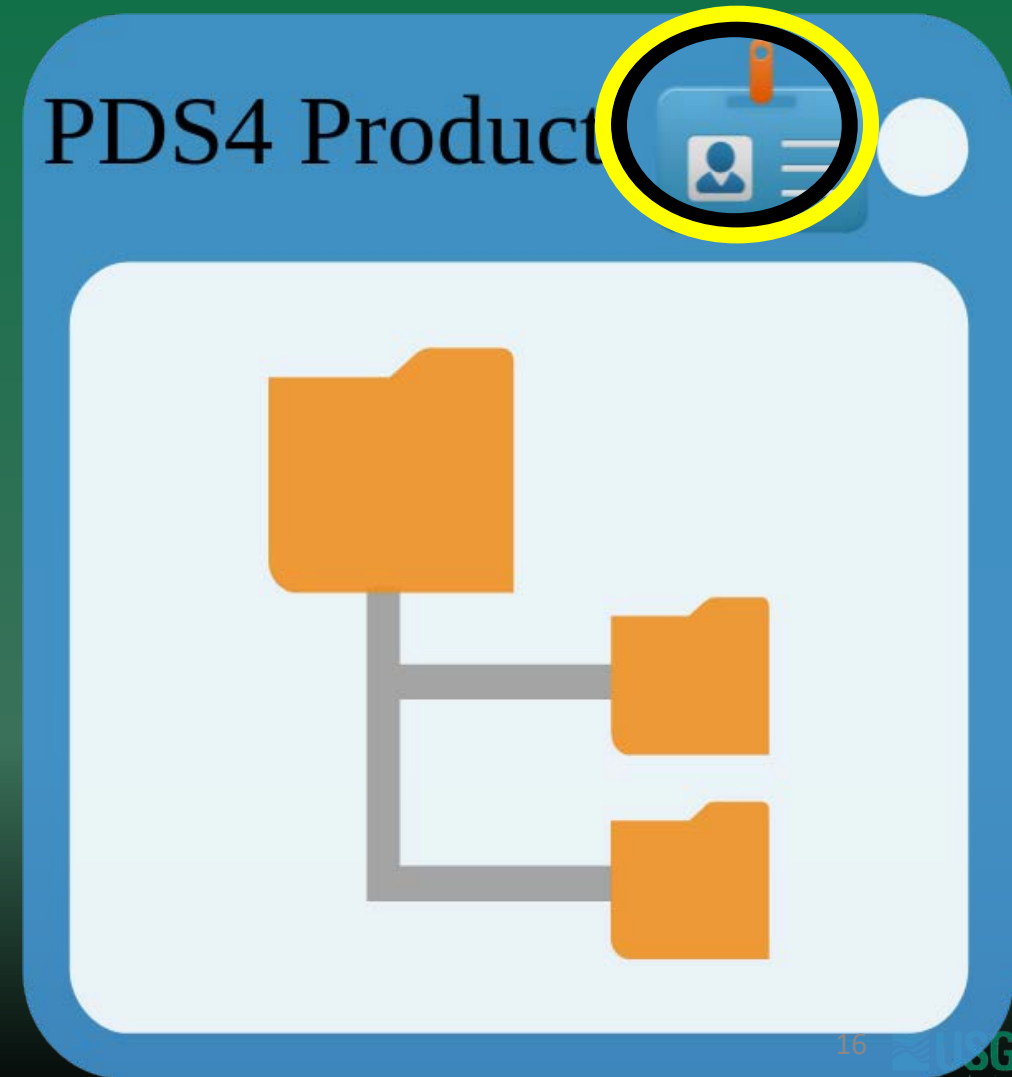
Product Bundle



Product Identity

`urn:nasa:pds:bundle:collection:product`

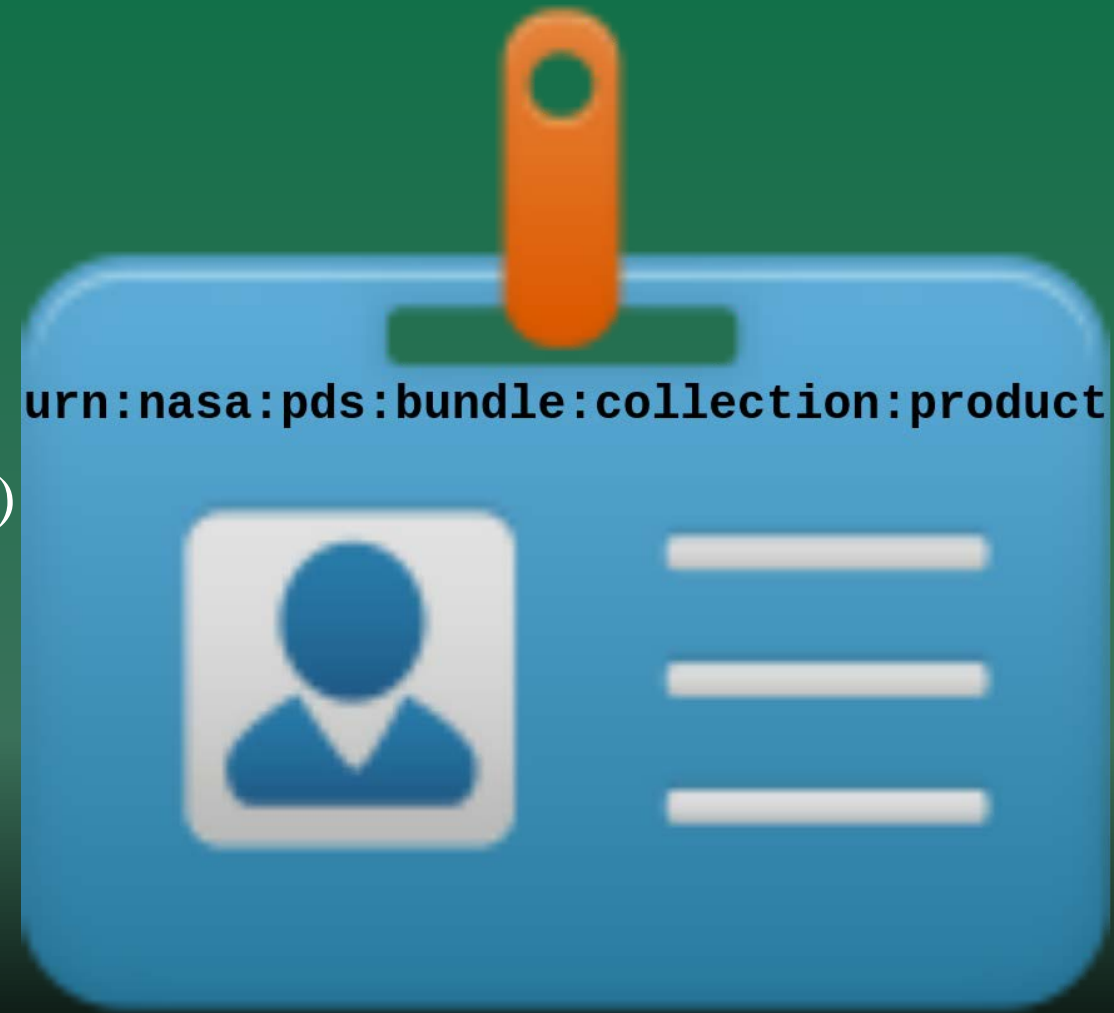
- Logical Identifiers (LIDs) unique across PDS
- Allowed Characters:
 - [a-z], [0-9], “-”, “.”, and “_”
 - Delimiters: “-” “.” “_”
 - Segments are delimited by colons
 - Max 255 characters
- Segments:
 - 1: URN identifier (static)
 - 2-3: nasa:pds (static)
 - 4: Bundle identifier
 - 5: Collection identifier
 - 6: Product identifier



LID Segments

urn:nasa:pds:bundle:collection:product

- **bundle ID:**
 - Shared by members of the bundle
 - Must be unique across all of PDS
- **collection ID:**
 - Shared by members of the collection
 - Begins with collection type (data, document, etc.)
 - Must be unique within the bundle
- **product ID:**
 - Must be unique within the collection



Archive Bundle LID

- The `bundle` product ID defines the `bundle` portion of the LID for its member collections
- The `collection` product ID defines the `collection` portion of the LID for its member basic products

urn:nasa:pds:**bundle**:collection:product



LIDVID: Logical IDentifier and Version ID

urn:nasa:pds:bundle:collection:product:version



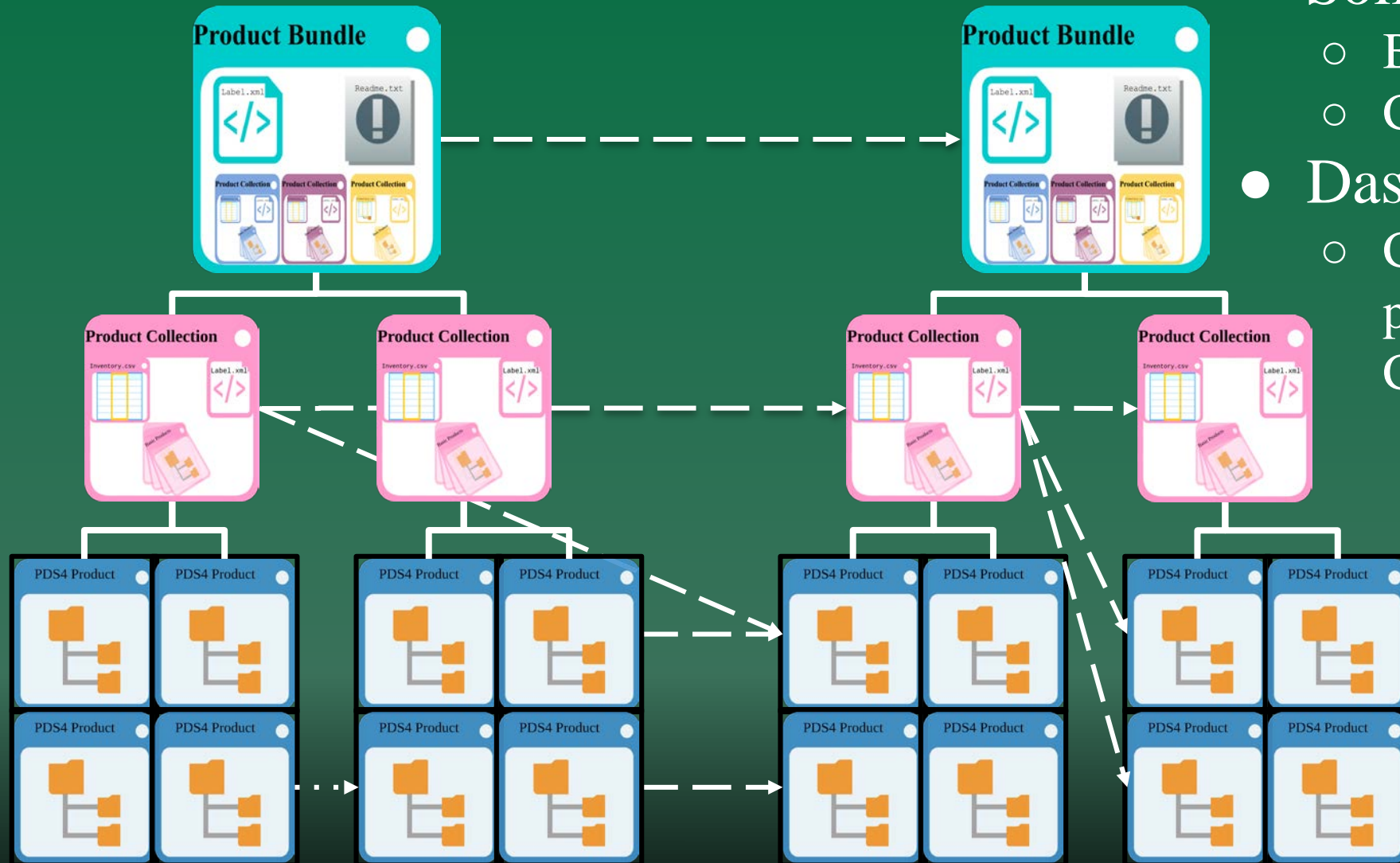
version 1



version 2
(reprocessed,
updated)

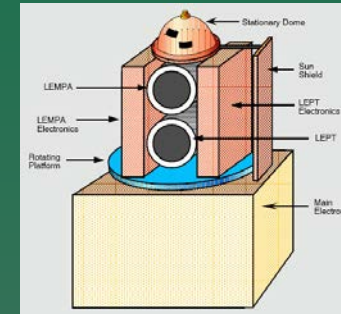
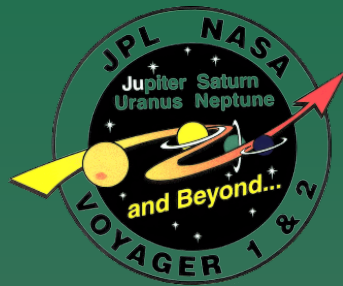
Referencing other Products

- Solid lines: Inventory
 - Bundle to Collection
 - Collection to Basic Products
- Dashed lines: References
 - Connections between products, possibly across Bundles and Collections



Context Products

- Physical or conceptual objects
 - Not physically part of the PDS archive
 - Institutions, missions, spacecraft, instruments, targets, etc.



- Provides
 - Association or provenance of other types of archive products
 - Under the stewardship of PDS Engineering Node
- **Not designed to be user documentation for those objects**

Archive Generation Procedure

Product planning and design should go from top down:



- collection inherits the bundle ID from the LID of their parent bundle. basic products inherit the bundle and collection IDs from parent bundle and collection

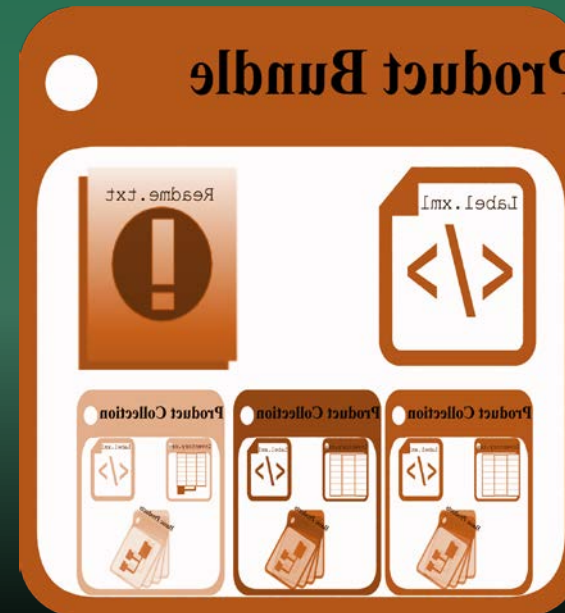
Product generation should go from bottom up:



- Metadata from basic products rolled into collection and bundle

Archive Design

- No absolute rule about PDS4 bundle organization
- Providers should consider:
 - What organization makes sense for the data?
 - What are other data providers on the project planning to do?
 - What are data users likely to find the most useful?
- **Consult with your curating node!**



PDS4 Data Dictionaries

- Data Dictionaries (*Presentation: Guinness*)
 - Reference for users, data producers
 - Ensures attributes and classes are used consistently and predictably
- Defines what may be used in a PDS4 product label
 - Human and s/w readable
 - Attributes:
 - Definitions, syntax, and semantic constraints on values
 - Classes:
 - Explicit list of defining attributes: required, optional, and repeatable
- 3 Kinds:
 - Common (PDS4 Data Dictionary)
 - Local Data Dictionaries
 - Discipline-specific
 - Mission-specific

Structure of a PDS4 Label

Identification Area		Context Area			Primary Result Summary		Investigation Area	
logical_identifier	title	comment	purpose	processing_..	description	name	type	
urn:nasa:pds:example.dph.sample_archive_bundle	PDS4 PPI VG2PLS Collect..	Observational Intent	Science	Raw	This is where a generic descript.. of the product goes.	VOYAGER	Mission	
version_id		Time_Coordinates			Science_Facets		Internal_Reference	
1.0		start_date_time 1979-07-06T00:00:42.687Z			wavelength.. domain discipline_n.. Gamma Ray Magnetos.. Particles		lidvid_referenc.. reference_ty.. urn:nasa:pds:context:investiga..:mission.voyager::1.0	
information_mod..		product_class		stop_date_time		Internal_Reference		
1.7.0.0		Product_Bundle		1979-07-09T23:59:06.436Z		urn:nasa:pds:context:investiga..:mission.voyager::1.0		
Citation_Information		Observing_System		Internal_Reference		Internal_Reference		
author_list publication_year DPH 2013		name Observing_System_Component		name Observing_System_Component		name Observing_System_Component		
editor_list description DPH		Obs.. name Internal_Reference VG2 urn:nasa:pds:context:instrument_host:spacecraft.vg2		name Internal_Reference PLS urn:nasa:pds:context:instrument:pls.vg2		name Internal_Reference urn:nasa:pds:context:instrument:pls.vg2		
Modification_History		reference_type is_instrument_host		reference_type is_instrument		reference_type is_instrument		
Modification_Detail		description instrument_host VG2		description instrument PLS		description instrument PLS		
modification_date description 2013-03-10 convers.. of PDS3 PPI archive to comply with PDS4 Informa.. Model		Target_Identification		Internal_Reference		Internal_Reference		
version_id		name type JUPITER Planet		lidvid_referenc.. reference_ty.. bundle_to_target		lidvid_referenc.. reference_ty.. bundle_to_target		
1.0								

File Area Text				Bundle Member Entry	
file_name	local_identi..	creation_da..	file_size	lid_reference	member_status
README.TXT	document.README.TXT	2013-11-11T11:59:04	12901	urn:nasa:pds:example.dph.sample_archive_bundle:browse	Primary
Stream_Text				reference_type	
name	local_identifier	parsing_sta..	description	bundle_has_browse_collection	
readme.txt	document.stream_text	7-Bit ASCII Text	offset 0	reference_type bundle_has_browse_collection	
Bundle_Member_Entry				Bundle_Member_Entry	
lid_reference	member_sta..	reference_ty..	lid_reference	lid_reference	lid_reference
urn:nasa:pds:example.dph.sample_archive_bundle:context	Primary	bundle_has_context_collection	urn:nasa:pds:example.dph.sample_archive_bundle:document	urn:nasa:pds:example.dph.sample_archive_bundle:xml_schema	urn:nasa:pds:example.dph.sample_archive_bundle:xml_schema
Bundle_Member_Entry				Bundle_Member_Entry	
lid_reference	member_sta..	reference_ty..	lid_reference	lid_reference	lid_reference
urn:nasa:pds:example.dph.sample_archive_bundle:data	Primary	bundle_has_data_collection	urn:nasa:pds:example.dph.sample_archive_bundle:document	urn:nasa:pds:example.dph.sample_archive_bundle:xml_schema	urn:nasa:pds:example.dph.sample_archive_bundle:xml_schema
Bundle_Member_Entry				Bundle_Member_Entry	
lid_reference	member_sta..	reference_ty..	lid_reference	lid_reference	lid_reference
urn:nasa:pds:example.dph.sample_archive_bundle:document	Primary	bundle_has_document_collection	urn:nasa:pds:example.dph.sample_archive_bundle:document	urn:nasa:pds:example.dph.sample_archive_bundle:xml_schema	urn:nasa:pds:example.dph.sample_archive_bundle:xml_schema

Anatomy of a PDS4 Label

Identification Area

Contains product identifying information

- LID & VID definition
- Authorship/citation information (optional)
- Product modification history (optional)

Observation/Context Area

Contains product provenance/background

- Observation time, target
- Scientific content description (science discipline, data processing level, wavelength range, etc.)
- Source (mission, observatory, instrument, etc.)
- Discipline-specific metadata (image display settings, geometry, etc.)
- Mission-specific metadata

XML Declaration

Product (Root) Tag

Identification Area

Observation/Context Area

Reference List

File Area

Anatomy of a PDS4 Label

Reference List

Contains links to other PDS4 products (by LID/LIDVID) and external publications

File Area

Contains a description of the labeled file

- File name
- File statistical information (optional: size, creation date, MD5 checksum)
- File format information
- Data file structural information
 - Array element descriptions
 - Table record and field descriptions

XML Declaration

Product (Root) Tag

Identification Area

Observation/Context Area

Reference List

File Area

Creating and Editing a PDS4 Label

- Any text editor can be used to create or edit a PDS4 label. However...
- An XML-aware editor will make the task much easier
- A schema-aware XML editor is even better
 - Helps guide you in what classes or attributes are allowed and where they occur in the label
 - Can do validation on the fly
- Schema-aware XML editors include
 - Oxygen (\$\$)
 - Eclipse (free): <http://www.eclipse.org/>
 - BaseX (free): <http://basex.org/home/>
- **Consult with your curating node!**

PDS4: Product Validation

- **Validate Tool**
 - Software designed to support validation of PDS4 product labels and product data
 - The associated specific schema for the product label specifies syntactic and semantic constraints
 - The product label itself specifies the constraints for the data
- A schema-aware XML editor can do validation on the fly
 - When current schema is loaded

PDS4 Tools

- PDS Label Assistant for Interactive Design (PLAID) (**DEMO: *Algermissen***)
 - Step by step GUI interface for building a PDS4 label that does not require experience with XML, PDS4 Schemas and Schematrons, or knowledge of the label requirements
 - Ensures that labels are valid, updated with PDS4 Schema
- AMMOS PDS Pipeline Service (APPS) (**Presentation: *Crombie***)
 - Mission pipeline tool for developing a PDS4 archive
- Local Data Dictionary Tool (LDDTool)
 - Software for parsing a local data dictionary definition file and generating PDS4 data standard files
- Generate Tool (**Presentation: *Cayanan***)
 - Software for generating PDS4 Labels using a user provided PDS4 XML template and input (source) data products
 - The PDS4 XML Templates use Apache Velocity variables and logic to generate the PDS4 labels
- Validate Tool (**Presentation: *Cayanan***)
 - Software for validating PDS4 product labels and product data. The associated specific schema for the product label specifies syntactic and semantic constraints
 - The product label itself specifies the constraints for the data
- Transform Tool (**Presentation: *Cayanan***)
 - Software for transforming PDS3 and PDS4 product labels and product data into common formats

PDS4: Data Processing Levels

PDS4 Data Processing Levels			
PDS4 Level	PDS4 Processing Level Description	CODMAC Level	NASA Level
Telemetry	An encoded byte stream used to transfer data from one or more instruments to temporary storage where the raw instrument data will be extracted. PDS does not archive telemetry data.	1	0
Raw	Original data from an instrument. If compression, reformatting, packetization, or other translation has been applied to facilitate data transmission or storage, those processes will be reversed so that the archived data are in a PDS-approved archive format. Often called Experiment Data Records (EDRs).	2	1A
Partially Processed	Data that have been processed beyond the raw stage but which have not yet reached calibrated status. These and more highly processed products are often called Reduced Data Records (RDRs).	2	1A
Calibrated	Data converted to physical units, which makes values independent of the instrument.	3	1B
Derived	Results that have been distilled from one or more calibrated data products (e.g., maps, gravity or magnetic fields, ring particle size distribution). Supplementary data, such as calibration tables or tables of viewing geometry, used to interpret observational data should also be classified as "derived" data if not easily matched to one of the other three above categories.	4+	2+
Sources:	-- https://pds.jpl.nasa.gov/policy/PolicyOnProcessingLevels03112013.pdf -- Nat'l Academy of Science Space Studies Board's Committee on Data Management and Computation (CODMAC) Edited Data (see National Academy press, 1986). -- http://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products/		

Changes in Archive Development?

- Minimal to no change (these are the same as from PDS3):
 - Milestones
 - Deliverables
 - Peer review procedures
 - Documentation requirements
- Mission data processing pipelines output PDS4 labels, not PDS3
- Design of data products may take a little longer
 - Fewer precedents to rely on (but this is changing rapidly!)
 - PDS4 standards continue to evolve (updates every 6 to 9 months), with older versions maintained
 - Validation of products for delivery to PDS should be easier

The selected Information Model for Europa Clipper mission will remain valid over the long term!

Essential Element

COMMUNICATION!

COMMUNICATION!

COMMUNICATION!

Make certain to identify the PDS Discipline Node that will be curating your archive early in the process and communicate with them regularly!

Resources for More Information

- PDS4 introduction
 - <https://pds.jpl.nasa.gov/datastandards/about/what.shtml>
- PDS4 concepts and documentation
 - <https://pds.jpl.nasa.gov/datastandards/documents/>
- SBN Wiki about PDS4
 - http://sbndev.astro.umd.edu/wiki/SBN_PDS4_Wiki
- PDS4 Schema
 - <https://pds.jpl.nasa.gov/datastandards/schema/>
- PDS4 Software and Tools
 - <https://pds.jpl.nasa.gov/tools/about/>
- PDS4 training materials
 - <https://pds.jpl.nasa.gov/datastandards/training/>

PDS Discipline Node Contacts

- **Atmospheres**

Lynn Neakrase

+1(575)646-1862

lneakras@nmsu.edu

- **Cartography and Imaging Sciences**

Lisa Gaddis

lgaddis@usgs.gov

- **Geosciences**

Ed Guinness

guinness@wunder.wustl.edu

Susie Slavney

slavney@wunder.wustl.edu

- **NAIF (SPICE)**

Boris Semenov

boris.semenov@jpl.nasa.gov

- **Planetary Plasma Interactions**

Joe Mafi

jmafi@igpp.ucla.edu

- **Ring-Moon Systems**

Mitch Gordon

mgordon@seti.org

- **Small Bodies**

Anne Raugh

raugh@astro.umd.edu

Backup

PDS4 Training Schedule for ECM

Europa Clipper PSG PDS4 Training Session Agenda

Friday June 15, 2018

Time (PDT)	Duration (min)	Topic	Speaker
8:00 AM	60	Introduction to PDS4 (40 min presentation, 20 min questions/discussion)	Moses Milazzo (USGS - PDS IMG Node)
9:00 AM	30	Discipline and Mission Dictionaries (20 min presentation, 10 min questions/discussion)	Ed Guinness (Washington University - PDS GEO Node)
9:30 AM	30	PDS4 Tools Overview (20 min presentation, 10 min questions/discussion)	Michael Cayanan (JPL - PDS Engineering Node)
10:00 AM	15	Break	
10:15 AM	30	Using PLAID for PDS4 Label Design (20 min presentation/demo, 10 min questions/discussion)	Stirling Algermissen (JPL - MGSS)
10:45 AM	30	APPS for PDS4 Bundle/Collection Development (20 min presentation, 10 min questions/discussion)	Kate Crombie (Indigo Information Services)
11:15 AM	30	Logistics when working with PDS mission POCs (20 min presentation, 10 min questions/discussion)	Moses Milazzo (USGS - PDS IMG Node)

WebEx Information

<https://jpl.webex.com/jpl/j.php?MTID=me5c5f3e8dd511e470a83a102ac5342f7>

Meeting number: 900 205 617

LID Bundle Identifier

urn:nasa:pds:bundle:collection:product

- Must be unique within PDS *mission-instrument[-description]*
 - *mission* = The mission ID
 - *instrument* = The instrument ID
 - *description* = A description (optional) to help to distinguish the bundle from others from the same mission and instrument
- Examples:
 - **ladee_nms**
 - **maven-swea-calibrated**

LID Collection Identifier

urn:nasa:pds:bundle:collection:product

- Must be unique within the bundle
- Starts with the collection_type value (lowercase)
- Collection identifiers typically take the form:

collection_type[-description], where

- *collection_type* = collection_type value (i.e. data, document, etc.)
- *description* = A description (optional) to help to distinguish the collection from others of the same type within the bundle (e.g. data type, mission phase, etc.)
- Examples:
 - **data**
 - **data_calibrated**
 - **data-svy-3d**

LID Product Identifier

urn:nasa:pds:bundle:collection:product

- Must be unique within the collection
- Typically consists of the base file name of the labeled file
- Examples:

nms_cal_hk__36127_20131203_104228

mvn_swe_l2_svy3d_20161208

- Design notes:
 - Uppercase characters must be converted to lowercase
 - File version numbers, and other variable portions of the file name should be omitted from the product identifier

Archive Schedule

Project Milestone	Phase	Activity	Status
	A/B	Identify Archive POCs	Complete
		Start DAWG Meetings	Complete
		Start Draft DMAP	Complete
		Start Draft ICDs	Initiated
Mission System PDR (6/2018)	B	Preliminary DMAP Ready	Complete
		Draft ICDs Ready	
Mission System CDR (11/2019)	C	Final DMAP Ready	
		Final ICDs Ready	
		Preliminary SISs Ready	
		Start Peer Reviews	
Launch -6 Months (12/2021)	D	Final SISs Ready	
		Peer Reviews Complete	
Data Received +6 Months	E	First Data Release	
Every 6 Months	E	Subsequent Data Releases	

DAWG Meeting Plans

- July/August 2018
 - Next DAWG meeting to discuss updated Archive Schedule and ICD status
- October/November 2018
 - Start periodic (bi-monthly) meetings through Phase C
- Proposed standing meeting day and time
 - Second Thursday of the month at 12:00 PM Pacific Time

Evaluation

Thank you for participating in our PDS4 training exercise. We would really appreciate your feedback on the quick survey below. Your answers are anonymous and are helpful to the development and improvement of our future training sessions.

<https://goo.gl/forms/KmoOkTZrv1pgSs2z2>



Thank you for your time!