



**Jet Propulsion Laboratory**  
California Institute of Technology

# PDS4 Development How-To: Labels, Bundles & Local Data Dictionaries

*Planetary Data Workshop*

*Flagstaff, AZ -- June 19, 2019*

Cristina M. De Cesare

[cristina.m.decesare@jpl.nasa.gov](mailto:cristina.m.decesare@jpl.nasa.gov)

Lead Mission Interface, PDS Cartography and Imaging Sciences Node

Jet Propulsion Laboratory, California Institute of Technology

# Outline

What is PDS? What is PDS4?

How do I develop PDS4 labels?

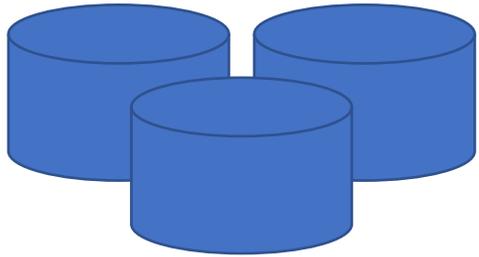
What are PDS4 dictionaries & which ones should I use?

How do I develop PDS4 bundles?

# What are PDS and PDS4?



The **Planetary Data System** (PDS) is NASA's repository for the distribution and long term preservation of NASA planetary data.



The **PDS Archive** is the digital data repository maintained by PDS.



The **PDS Standard** are requirements and constraints designed to ensure the usability of data in the PDS Archive throughout the lifetime of the archive.

**PDS4** is the latest version of the PDS Standard. (PDS4 is **not** a data format.)

# PDS4 Archive Organization

- There are 3 primary types of products in PDS4:
- **Basic Products** are the smallest unit of a PDS4 archive. They consist of an individual label and the associated file or files.
- Related basic products of the same type may be grouped together into a **Collection**.
- Related collections may be grouped together into a **Bundle**.

Basic  
Products



Collection  
Products

Bundle  
Products

# What are PDS4 labels?

- **PDS4 labels** are XML files that describe the contents of one or more files:
  - Observational data
  - Documentation files
  - Calibration data
- A PDS label, along with the file(s) that it describes, constitute a **PDS Product**.

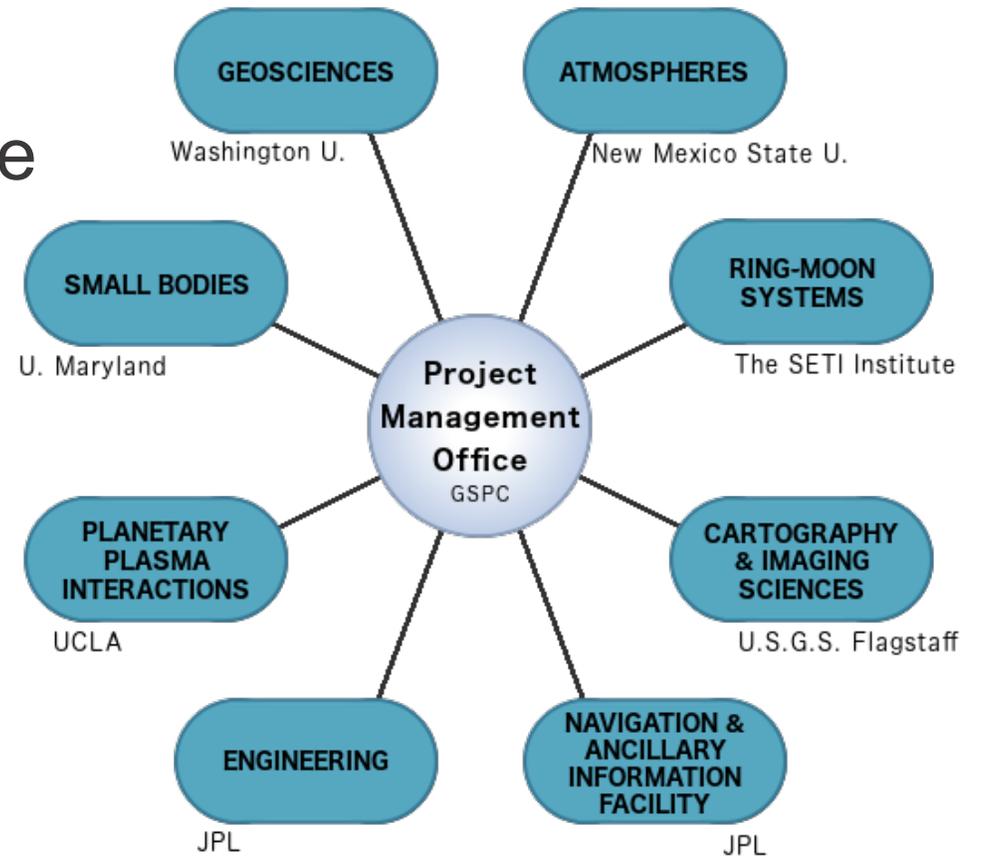
## PDS Product



# I need PDS4 labels. Where do I start?

## Ask yourself these questions:

- What kinds of data products will I be archiving?
- Are there many different types of data products?
  - Do these types each have unique metadata?
- What PDS Node will this data be archived with?



# How do I create my 1<sup>st</sup> label?

1. Choose 1 data file to start with.
2. Install a text editor or XML editor
  1. Notepad++
  2. OxygenXML
3. Decide which PDS tool(s) to use:
  1. **PLAID** (PDS Label Assistant for Interactive Design) is a good starting point. It's an interactive web-based tool that will help you build a PDS4 label step-by-step.
  2. Is your data file a CSV or fixed-width table? Will your archive contain many such files?
    1. Try **OLAF**, a web application that will help you build a complete PDS4 archive bundle from the data files you upload.
4. Build your label, until you have a rough draft.

# Anatomy of a PDS4 Label

## XML Declaration

- XML identification tag
- Schematron location information

## Product Tag

- Root product type tag
- Namespace declarations
- Schema location information

## *XML Declaration*

## *Product (Root) Tag*

### *Identification Area*

### *Observation/Context Area*

### *Reference List*

### *File Area*

# 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
- Scientific content description (science discipline, data processing level, wavelength range, etc.)
- Target
- 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*

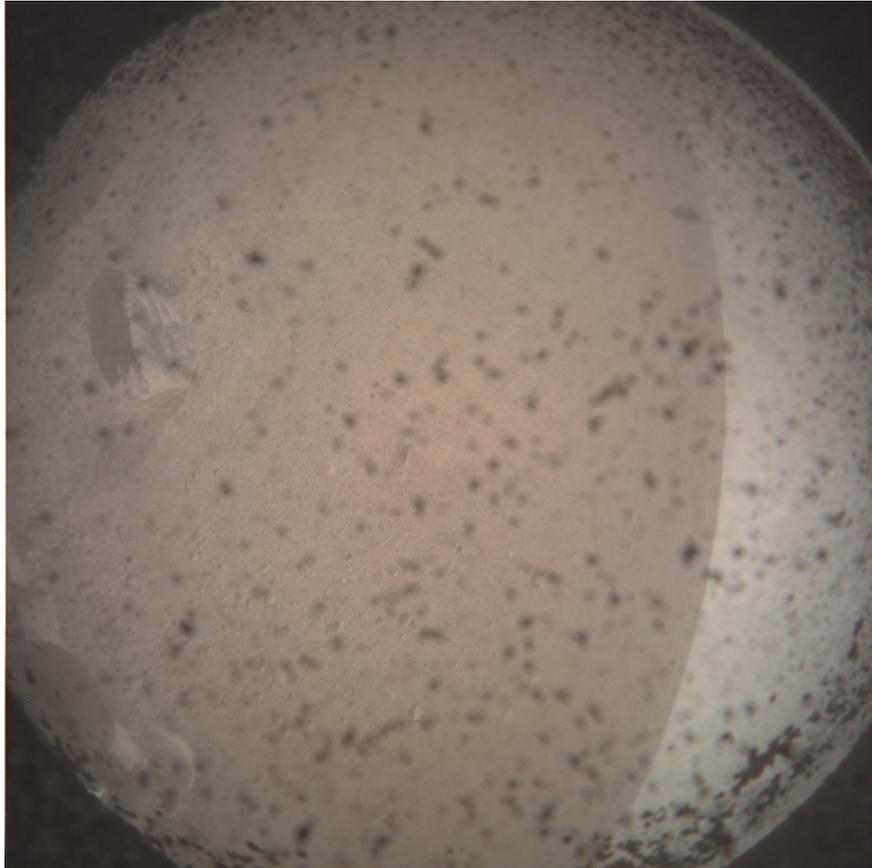
*Identification Area*

*Observation/Context Area*

*Reference List*

*File Area*

# InSight case study: example label metadata



```
<img:Exposure>  
  <img:exposure_count>2</img:exposure_count>  
  <img:exposure_duration unit="ms">329.13</img:exposure_duration>  
  <img:exposure_duration_count>53</img:exposure_duration_count>  
</img:Exposure>
```

```
<msn:mission_phase_name>SURFACE MISSION</msn:mission_phase_name>  
<msn:product_type_name>EDR</msn:product_type_name>  
<msn:spacecraft_clock_start>596533559.23421</msn:spacecraft_clock_start>  
<msn:spacecraft_clock_stop>596533559.44990</msn:spacecraft_clock_stop>  
<msn:spacecraft_clock_partition>1</msn:spacecraft_clock_partition>
```

Observational data file  
raw image from InSight sol 0

# How do I validate my PDS4 label?

- Use PDS4 Validate Tool
- If using Oxygen XML Editor (or other XML editor), you can also validate your label from inside the software tool.

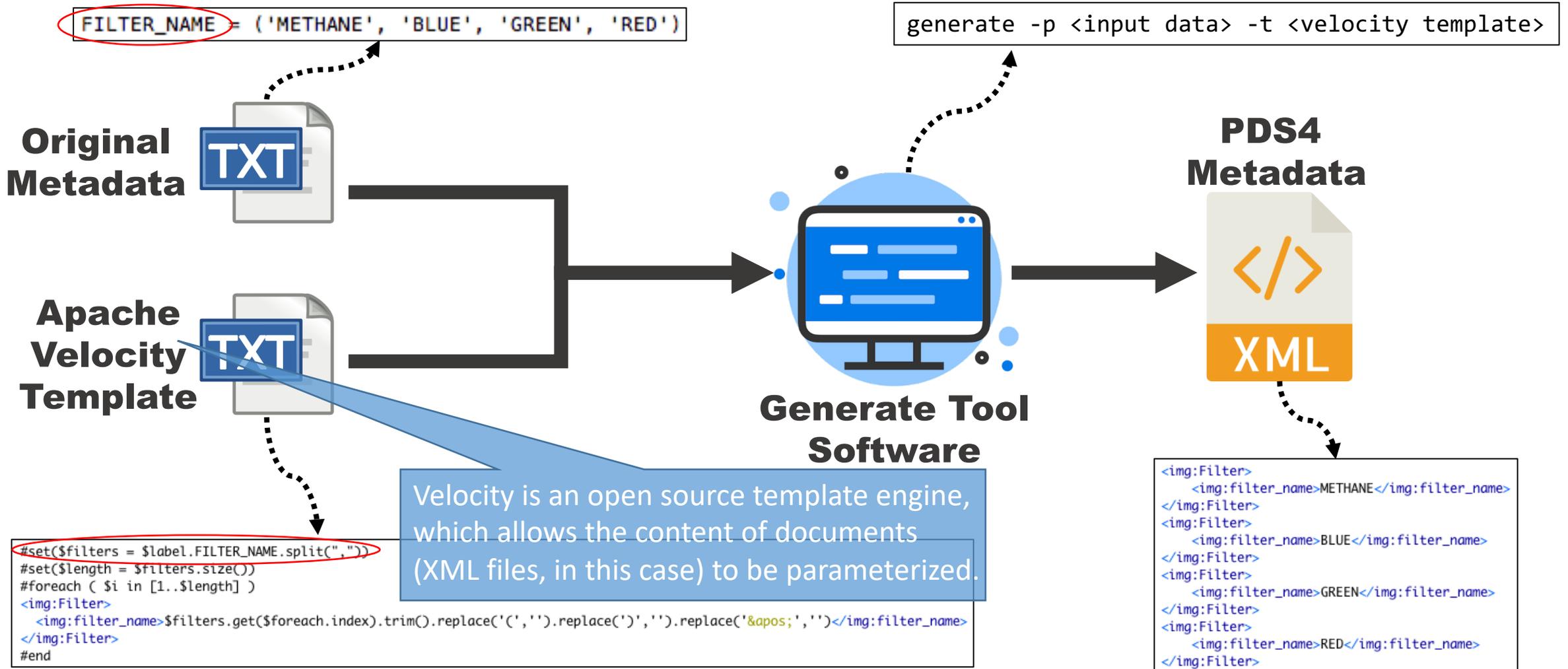
# What if my label doesn't validate?

- Don't panic!
- The validation report will show you a list of errors.
  - Don't forget, in PDS4, ordering matters.
- If you're stuck, contact your PDS Node representative for help.

# I have 1 valid label. Now what?

- Use this label as a model when you develop labels for all the other data files of this type.
- You can use the PDS Generate Tool to mass-produce labels:
  1. Make a copy of your valid label, and replace specific metadata values with variables.
  2. Run Generate to produce labels for all your data files.
- Or, you can write your own script to do this, in a language of your choice.

# Using Generate to produce labels



# Remember this?

## Observation/Context Area

Contains product provenance/background

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

## Anatomy of a PDS4 Label

*XML Declaration*

*Product (Root) Tag*

*Identification Area*

*Observation/Context Area*

*Reference List*

*File Area*

# What are PDS4 data dictionaries?

- PDS4 data dictionaries...
  - Define the meaning and structure of XML classes & attributes to be used in PDS4 label files.
  - Specify XML tags, their meanings, and acceptable values
  - Include an XML Schema file (\*.xsd) & XML Schematron file (\*.sch)
- 3 categories of dictionary:
  - PDS core
  - Discipline dictionary
  - Mission dictionary
- The latest PDS4 dictionaries are published on the PDS website:
  - <https://pds.nasa.gov/datastandards/schema/released/>

# 3 categories of data dictionary

## PDS Information Model

PDS  
common  
dictionary

## Local Data Dictionaries (LDDs)

### Discipline Dictionaries

Cartography

Display

Geometry

Imaging

Processing Info.

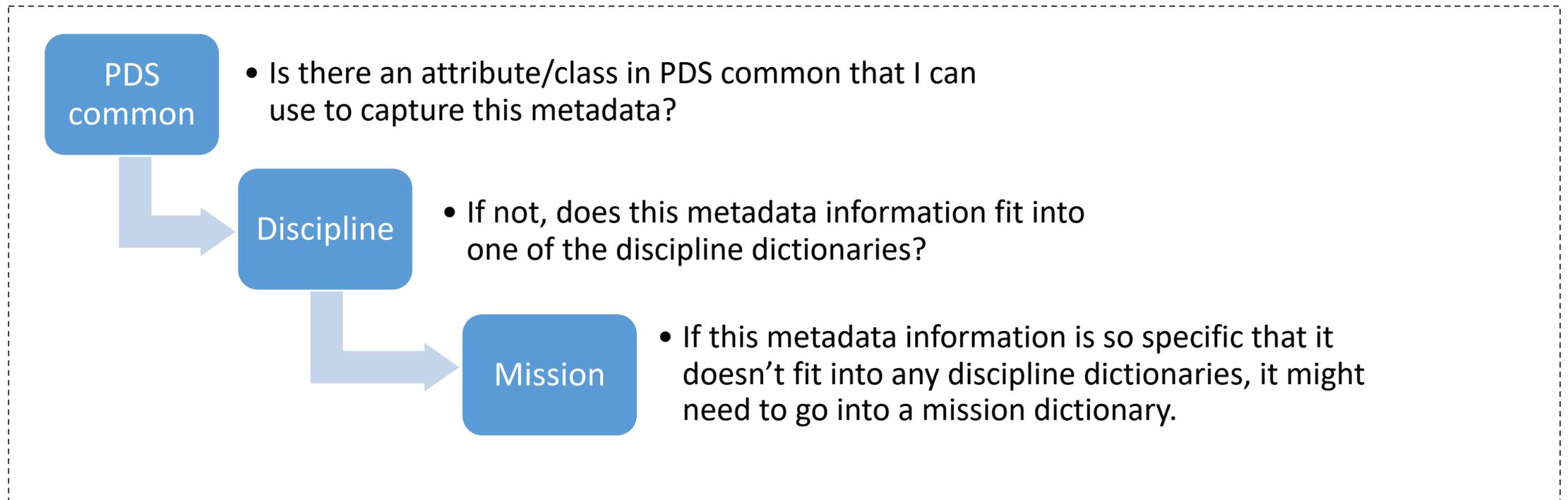
Ring-Moon Systems

Spectral

Mission  
Dictionaries

# Which dictionaries should I use?

- Every PDS4 label must use the PDS common dictionary.
  - It defines the essential, required components of even the simplest label.
- Discipline & mission dictionaries are optional, and their use will be driven by the types of metadata you wish to include in your labels.



# What if my metadata doesn't fit into any dictionary?

- Ask your PDS Node representative for help.
- They can advise you on...
  - Whether any existing data dictionaries could be updated in order to support your metadata.
    - And what this process looks like
  - Whether you ought to create your own mission-specific local data dictionary
    - And how to do so.
    - Refer to Anne Raugh's upcoming talk about building Local Data Dictionaries!

# **I've got my labels & dictionaries. How do I put them all together?**

- Collections
  - Products are grouped into 'collections' by type and content.
  - Examples:
    - Data collection
    - Document collection
    - Supplementary collection
- Bundle
  - A list of collections
  - Some missions may have 1 bundle, while others may have multiple

# How do I design my archive?

There is no hard and fast rule governing how a PDS4 archive is to be organized.

Data providers may want to consider the following questions:

- 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!



# InSight case study: archive design

- insight\_cameras
  - browse
  - data
    - sol
      - 0001
        - mipl
          - edr
            - icc
            - idc
          - rdr
            - icc
            - idc
            - mesh
            - mosaic
- document
- miscellaneous
- xml\_schema

# How do I develop PDS4 bundles?

- Write your own scripts to:
  - Generate labels into their respective collection directories
  - Place those collection directories into a bundle directory
  - Generate collection-level and bundle-level metadata files
  - Invoke the PDS4 Validate Tool on the new bundle
- Such scripts can be integrated into existing data processing pipelines for a mission/observation.
- Or, consider using APPS (AMMOS-PDS Pipeline Service):
  - Automates the collection of data products for PDS4 bundle generation

# References

1. NASA PDS. (2018) *What is PDS4?* Retrieved from <https://pds.jpl.nasa.gov/pds4/about/what.shtml> .
2. NASA PDS. (2018) *PDS Concepts*. Retrieved from [https://pds.jpl.nasa.gov/datastandards/documents/concepts/Concepts\\_1.11.0.pdf](https://pds.jpl.nasa.gov/datastandards/documents/concepts/Concepts_1.11.0.pdf)
3. Moses Milazzo. (2018) *Introduction to PDS4*. Retrieved from <https://pds.nasa.gov/datastandards/training/2018-ec/ec-pds4-introduction-201806.pdf>
4. NASA PDS. (2017) *PDS4 Training Exercise*. Retrieved from <https://pds.nasa.gov/datastandards/training/2017-agu/pds4-training-agu-2017-v2.pptx>
5. Jet Propulsion Laboratory. (2018) *InSight ICC EDR Observational Product - c000m0000\_596533559edr\_f0000\_0106m*. Retrieved from [https://pds-imaging.jpl.nasa.gov/data/nsyt/insight\\_cameras/data/sol/0000/mipl/edr/icc/C000M0000\\_596533559EDR\\_F0000\\_0106M9.xml](https://pds-imaging.jpl.nasa.gov/data/nsyt/insight_cameras/data/sol/0000/mipl/edr/icc/C000M0000_596533559EDR_F0000_0106M9.xml)
6. “Text-xml” by RRZEicons is licensed under CC BY 2.0. <https://commons.wikimedia.org/wiki/File:Text-xml.svg>.
7. “Text-txt” by RRZEicons is licensed under CC BY 2.0. <https://commons.wikimedia.org/wiki/File:Text-txt.svg>.

**Questions?**



**Jet Propulsion Laboratory**  
California Institute of Technology

---

[jpl.nasa.gov](http://jpl.nasa.gov)