

PDS4 DEVELOPMENT HOW-TO: LABELS, BUNDLES, AND LOCAL DATA DICTIONARIES. C. M. De Cesare¹ and J. H. Padams¹, ¹Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, 91109-8099, USA.

Introduction: 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 the PDS. The PDS Standard is the set of requirements and constraints designed to ensure the usability of data in the PDS Archive throughout its lifetime. The PDS Standard is derived from the PDS4 Information Model [1], a formal model that consistently defines the PDS4 products, from their allowed formats to the associated metadata.

PDS4 is the latest version of the PDS Standard. Every data product in the PDS Archive is accompanied by a PDS label file, which contains metadata. In PDS4, metadata labels are represented using XML (Extensible Markup Language). XML is an international standard that provides both syntax and structure for describing data. The adoption of XML allows for increased consistency of metadata, which, in turn, improves accessibility and usability of the archived data.

In order to archive data with the PDS, data providers must produce PDS4 metadata labels for their data products, assemble both label files and product files into a PDS Bundle, and deliver this Bundle to the appropriate PDS Discipline Node [2]. This process often includes development of a Local Data Dictionary (LDD), when the data provider would like to define a new PDS4 dictionary to contain metadata specific to their mission or investigation.

To a newcomer, this process may be confusing, and the only way to learn it is by doing. In an effort to make this a less daunting task for data providers and PDS personnel alike, I will describe some best practices for PDS4 development, based on experiences with the InSight, Juno, and Mars Pathfinder missions.

Label Development: The process of archiving data with the PDS usually begins with PDS4 label development. You have data products that you want to archive, and you need to make labels for each one of them. In order to do this, you must first ask yourself some questions about the data:

- What kinds of data products will be archived?
- Which PDS Node will this data be archived with?
- Are there many different types of data products?
 - If so, will the labels for each data product contain very different metadata?

From here, you will need to design your PDS4 labels and, especially if you have numerous data products, develop a process for systematically producing labels in bulk. The PDS offers a variety of tools to assist you with this:

- PLAID (PDS Label Assistant for Interactive Design) [3]: An interactive web-based tool which helps you build a PDS4 label step-by-step.
- Generate Tool: Software for generating PDS4 labels using input data products and an Apache Velocity [4] template of a PDS4 XML label. This tool is especially useful if you need to migrate existing PDS3 metadata to PDS4 [5].
- OLAF [6]: A web application that allows individuals to submit source data files and produces the data set as a complete PDS4-compliant archive package, ready for peer review and archiving. OLAF currently accepts CSV (comma-separated values), fixed-width tables, 2D images, and documents.

Local Data Dictionaries: Every attribute and class that is used in any PDS label must first be defined in a data dictionary. A data dictionary defines the attributes and classes that may be used in PDS4 product labels. Data dictionaries are classified as either common, discipline-specific, or mission-specific [2]. While developing your PDS4 labels, you may find that you have metadata that (1) is important enough to be included in your PDS4 labels and (2) doesn’t fit into any existing element in any of the PDS4 dictionaries. In this case, you may want to develop a “mission dictionary”, which will contain the metadata that is specific to your investigation and thus is not generic enough to be included in the PDS common model or Discipline Dictionaries.

When trying to determine how to represent your metadata in PDS4 labels, I recommend the following approach:

1. Check PDS common model [1] for classes/attributes that fit your metadata.
2. Check relevant Discipline Dictionary [7] (such as Geometry or Imaging) for classes/attributes that fit your metadata.
3. If you still haven’t found a good location to put your metadata, consult your PDS Node representative to see if they have a suggestion.
4. Otherwise, this metadata truly doesn’t fit anywhere else and should go into a mission-specific

dictionary. Ask your Node representative for guidance on creating this dictionary. (Not sure who your Node representative is? Email the PDS Operator at pds_operator@jpl.nasa.gov.)

Bundle Development: Once you have developed your PDS4 labels and written your mission-specific dictionary (if needed), you're ready to validate your labels and assemble them, alongside their data products, into a PDS4 bundle. The *PDS4 Data Provider's Handbook* (DPH) [2] offers a detailed explanation of what a PDS4 bundle must contain and how it should be structured. If you have many data products and labels, you may wish to write a script to produce your bundle, according to the structure described in the DPH. When in doubt, consult with your Node representative for guidance on bundle creation.

To check the validity of your PDS4 labels, the PDS offers the Validate Tool [8]. The Validate Tool is a simple command-line utility that will check your PDS4 bundle, and the labels within it, to ensure that the bundle follows PDS4 standards and maintains referential integrity. The Validate Tool will output a log file containing details of any errors.

Conclusion: The PDS is NASA's archive for planetary data, and we want to demystify the process of submitting data to our archive. We have a variety of resources available to assist data providers with PDS4 development, and we will continue to build real-life examples for users to draw upon.

References:

- [1] NASA PDS (2019) *Information Model Specification*. Retrieved from <https://pds.nasa.gov/datastandards/documents/im>
- [2] NASA PDS (2018) *The PDS4 Data Provider's Handbook*. Retrieved from https://pds.jpl.nasa.gov/datastandards/documents/dph/current/PDS4_DataProvidersHandbook_1.11.0.pdf
- [3] Jet Propulsion Laboratory, California Institute of Technology (2019) *PLAID*. Retrieved from <https://plaid.jpl.nasa.gov>
- [4] The Apache Software Foundation (2016) *The Apache Velocity Project*. Retrieved from <https://velocity.apache.org>
- [5] NASA PDS (2019) *Generate Tool*. Retrieved from <https://pds.nasa.gov/tools/about/generate>
- [6] PDS Small Bodies Node (2019) *OLAF On-Line Archiving Facility*. Retrieved from <https://sbnapps.psi.edu/olaf-node>
- [7] NASA PDS (2019) *Released Schemas*. Retrieved from <https://pds.nasa.gov/datastandards/schema/release>

- [8] NASA PDS (2019) *Validate Tool*. Retrieved from <https://pds.nasa.gov/tools/about/validate>