



Jet Propulsion Laboratory
California Institute of Technology

Bridging Archival Standards: Building Software to Translate Metadata Between PDS3 & PDS4

Planetary Science Informatics and Data Analytics Conference

St. Louis, MO -- April 25, 2018

Cristina M. De Cesare
cristina.m.decesare@jpl.nasa.gov
Lead Mission Interface, PDS Imaging Node
Jet Propulsion Laboratory, California Institute of Technology

Outline

Intro to PDS & PDS4

Challenges of translating between PDS3 & PDS4

A Solution: The Label Mapping Tool

What it does & how it works

Future software development

Intro to PDS & PDS4

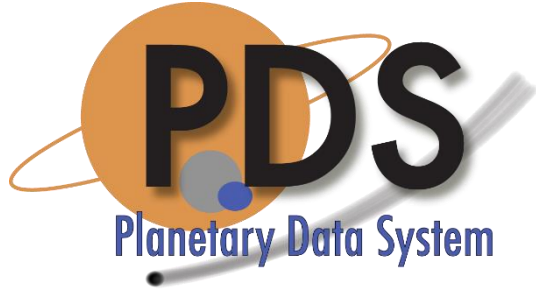
Challenges of translating between PDS3 & PDS4

A Solution: The Label Mapping Tool

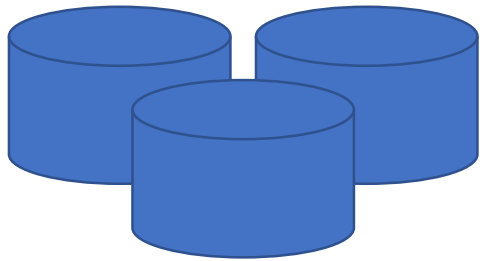
What it does & how it works

Future software development

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.)

PDS Labels

PDS Product



- A file containing PDS metadata is called a **PDS Label**.
- A PDS label, along with the file or files that it describes, constitute a **PDS Product**.

Intro to PDS & PDS4



Challenges of translating between PDS3 & PDS4

A Solution: The Label Mapping Tool

What it does & how it works

Future software development

Labels in PDS3 vs. PDS4

Old Standard: PDS3	New standard: PDS4
Object Description Language (ODL) text file 	XML (Extensible Markup Language) file 
“keyword = value” syntax, inconsistent document structure	Standardized syntax & structure, defined by PDS Information Model (IM)
Difficult to enforce standards & best practices	XML validation provides enforcement of complex data relationships, rules & standards.
Inconsistent implementation → Negative impact on usability & interoperability	Improved metadata consistency → Improved accessibility & usability of archived data

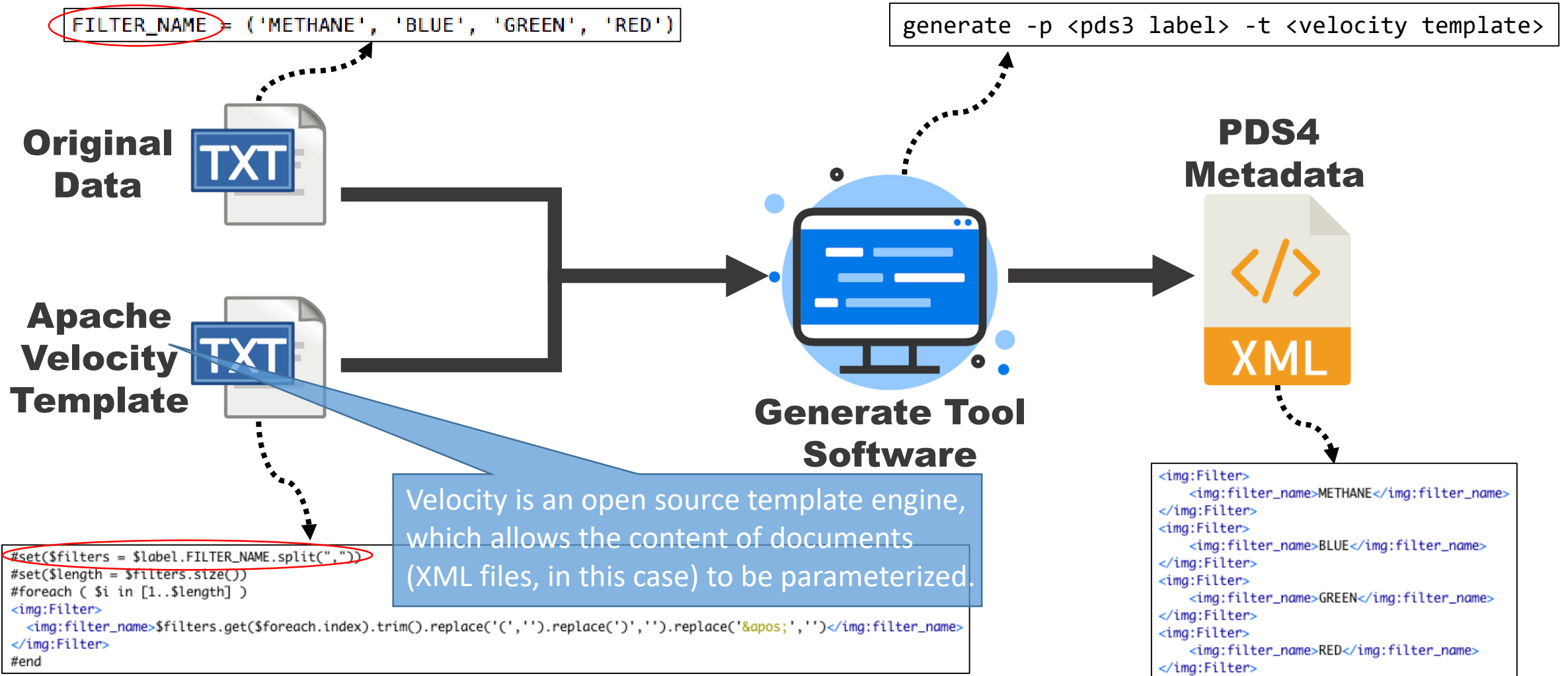
PDS3 Label Example

```
OBJECT                = IMAGE
  LINES               = 2048
  LINE_SAMPLES        = 1648
  SAMPLE_TYPE         = UNSIGNED_INTEGER
  LINE_PREFIX_BYTES   = 0
  LINE_SUFFIX_BYTES   = 0
  SAMPLE_BITS         = 8
  SAMPLE_BIT_MASK     = 2#11111111#
  MD5_CHECKSUM        = "80f6b30f5686b7e8c60032ab9044d91f"
END_OBJECT
```


PDS4 Label Example

```
<File_Area_Observational>
  <File>
    <file_name>JNCE_2011238_00A00002_V01.IMG</file_name>
    <local_identifier>DATA_FILE</local_identifier>
    <creation_date_time>2016-10-27T14:36:37</creation_date_time>
    <file_size unit="byte">2520</file_size>
    <md5_checksum>80f6b30f5686b7e8c60032ab9044d91f</md5_checksum>
  </File>
  <Array_2D_Image>
    <local_identifier>jnce_2011238_00a00002_v01</local_identifier>
    <offset unit="byte">0</offset>
    <axes>2</axes>
    <axis_index_order>Last Index Fastest</axis_index_order>
    <Element_Array>
      <data_type>UnsignedMSB8</data_type>
      <unit>DN</unit>
    </Element_Array>
    <Axis_Array>
      <axis_name>Line</axis_name>
      <elements>2048</elements>
      <sequence_number>1</sequence_number>
    </Axis_Array>
    <Axis_Array>
      <axis_name>Sample</axis_name>
      <elements>1648</elements>
      <sequence_number>2</sequence_number>
    </Axis_Array>
  </Array_2D_Image>
</File_Area_Observational>
```

Translation between PDS3 & PDS4



The Problem

PDS3 → PDS4 translation is tedious & requires knowledge of the data!

But datasets need to be converted from PDS3 to PDS4.

- Legacy missions already archived in the PDS:
 - Mars Pathfinder
 - Phoenix
- Ongoing & upcoming missions still producing PDS3 ODL labels in their ground data processing pipelines:
 - MER
 - MSL
 - InSight
 - Mars 2020

Learning curve for PDS4 label development

- Small data providers lack personnel.
- Big “flagship” missions have a large volume of data, more complicated labels, etc.

Intro to PDS & PDS4

Challenges of translating between PDS3 & PDS4

A Solution: The Label Mapping Tool

What it does & how it works

Future software development

A Software Solution: PDS Label Mapping Tool (LMT)

New Python software developed by PDS IMG Node at JPL

Input:

- PDS3 ODL label file
- Velocity template file

An XPath is a reference to a specific location in an XML document.

Output:

- CSV file -- contains mappings between PDS3 keywords and equivalent PDS4 XPath

PDS3 Vicar Group & Keyword	XPath
IMAGE.LINES	pds:File_Area_Observational/pds:Array_2D_Image/pds:Axis_Array[1]/pds:elements
IMAGE.LINE_SAMPLES	pds:File_Area_Observational/pds:Array_2D_Image/pds:Axis_Array[2]/pds:elements
IMAGE.MD5_CHECKSUM	pds:File_Area_Observational/pds:File/pds:md5_checksum

Label Mapping Tool

What does it do?

Aids developers performing PDS3 to PDS4 archive conversions.

- Helps ensure that all keywords from the original metadata are captured in the PDS4 label.

Provides mappings where needed:

- Mission Software Interface Specification (SIS) documents
- PDS Label Assistant for Interactive Design (PLAID) software
- PDS Image Atlas web search

How does it work?

- Leverages the effort already invested to develop Velocity Template for a mission/instrument.
- Parses PDS3 label to extract keywords.
- Parses Velocity Template using *lxml* Python library.
- Iterates over XML elements, looking for references to variables that match the PDS3 keywords.
- Pairs matching keywords & XPath's into a CSV.

Label Mapping Tool: the Future

Open source release

Develop new pipeline to take in LMT output and:

- Generate Property Maps
- Generate Terminological Entry data
- Build database of mappings from PDS3 keywords to:
 - PDS4 XPath
 - Mission-specific nuances for attribute/class definitions
 - VICAR keyword/group
 - Etc.
- Provide PDS4 JSON snippets via web service

PDS Resources

Tools

- Generate Tool
 - <https://pds.jpl.nasa.gov/pds4/software/generate/>
 - For more info, stop by my poster in this afternoon's session.
- PLAID
 - <https://plaid.jpl.nasa.gov/>

Documentation

- PDS4 Training
 - <https://pds.jpl.nasa.gov/pds4/training/>
- Data Provider's Handbook
 - <https://pds.jpl.nasa.gov/pds4/doc/dph/>

Credits & References

Credits

- Jordan Padams, for training me on all things PDS.
- Steve Hughes & Bob Deen, for helping to design the future of LMT.
- Joe Mafi, for the PDS intro slides.

References

1. NASA Jet Propulsion Laboratory. (2018) *What is PDS4?* Retrieved from <https://pds.jpl.nasa.gov/pds4/about/what.shtml> .
2. “Text-xml” by RRZEicons is licensed under CC BY 2.0. <https://commons.wikimedia.org/wiki/File:Text-xml.svg>.
3. “Text-txt” by RRZEicons is licensed under CC BY 2.0. <https://commons.wikimedia.org/wiki/File:Text-xml.svg>.
4. The Apache Software Foundation. (2016) *The Apache Velocity Project*. Retrieved from <http://velocity.apache.org>.



Jet Propulsion Laboratory
California Institute of Technology

jpl.nasa.gov