

MIGRATING DATA, LOCAL DICTIONARY USE, LEVERAGING PYTHON

L.F. Huber, V.I. Martinez, A. Hyder, T. Güth, Z. Stevenson, J. Karl, J.J. Johnson, N. Chanover, L.D.V. Neakrase
Planetary Data System, Atmospheres Node, Dept. of Astronomy, P.O. Box 30001, MSC 4500, New Mexico State
University, Las Cruces, NM 88003-8001

Introduction. The PDS Atmospheres Node is currently in the process of migrating its archived data holdings from the PDS3 to PDS4 archiving standard. The process encompasses a large number of missions and data types and processing levels dating back to the late 1960s through present missions. The goal of this work is to bring all data holdings up to the PDS4 standard, allowing for better integration across all investigations.

PDS3-PDS4 and Python. The Atmospheres Node has developed a collection of Python routines and scripts to make the migration process as automatic as possible, even when working with more complex labels, such as those used by several of the *Juno* instruments. The Python codes handle reading the valid PDS3 labels (.lbl) using translation files to convert to equivalent PDS4 attributes. The resultant PDS4 output creates .xml labels for all the pieces of the bundle utilizing the various parts of the PDS3 structures. The archive bundles are created all at once through the code and PDS4 archive bundles are built with minimal human effort. Resultant bundles are then validated against the PDS4 standard and released alongside the certified PDS3 versions of the same data.

Initial testing of the code was done on *Phoenix Lander* atmospheric data in the earlier days of PDS4 development and has evolved over time alongside the PDS4 Information Model (IM) as refinements are released. Current codes convert PDS3 to PDS4 up to IM Version 1.11.0.0 [at the time of this abstract]. Current mission migration efforts are handling atmospheric data from *Mars Global Surveyor*, *Mars Odyssey*, *Mars Reconnaissance Orbiter*, *MESSENGER*, *Mars Science Laboratory*, *Juno*, *Cassini*, *Huygens* and *Viking*.

With more complex datasets and the inclusion of the larger mission data, there has been a need to include local data dictionaries. Local data dictionaries may incorporate mission attributes in a mission dictionary or discipline specific attributes (e.g., display, geometry, cartography, spectroscopy, etc.) that are defined outside the core PDS4 IM. Dictionary support within PDS4 labels is found in specific places and the core migration codes have required the inclusion of local data dictionaries to complete migration of several datasets.

Features and Capabilities. Several different Python scripts and other supporting files work together to properly execute a migration. A blank XML template that corresponds to the overall structure of the PDS3 labels is the starting point. For datasets containing Local Data Dictionaries, the XML declaration at the top of this template must be modified to include the appropriate targetNamespace and schemaLocation for those dictionaries in order for those tags to show up in the template. A keyword lookup for translating keywords from PDS3 to PDS4 is used to fill in many of the repeated tags of the label, the LID's and/or LIDVID's, and points to the tags used within the Local Data Dictionaries. A compilation of student-written Python methods (PDS_Module) capable of handling different tables, file types, and tag-filling methods is continuously being updated/modified. One main Python script is generally used per dataset to read through the PDS3 label, call the respective functions from the Module, and output the corresponding PDS4 .xml files.

The amount of human effort that goes into preparing a migration primarily depends on the datasets themselves. The more complex datasets (e.g., FITS files, files containing SPICE kernels, or datasets with multiple instruments) require more effort as the structure of these files is rather complicated and the methods used to handle these are more complex. As more migrations are done, this process becomes more intuitive. Errors become easier to work through and show up less often. Files other than data files (e.g., documents) take time to complete as most of them are done by hand rather than by code.

Extensibility and Future Plans. The Atmospheres Node plans to completely migrate its holdings in the next 5 years. This will also entail maintaining the migration code for continuing missions such as *Juno*, *Mars Reconnaissance Orbiter* and *Mars Science Laboratory* until those missions have ceased science operations.

As the PDS4 Information Model continues to mature, it is likely that we will “re-migrate” our holdings in order to make better use of improvements to the IM. Some of the more complex data sets will be reserved until the end of our efforts.