STATUS OF GEOSCIENCES NODE'S MIGRATION OF LEGACY DATA TO THE PDS4 STANDARD.

E.A. Guinness, J.G. Ward, D.V. Politte, T.C. Stein, P.K. Byrne, McDonnell Center for Space Sciences, Dept. of Earth and Planetary Sciences, Washington University in St. Louis, 1 Brookings Drive, CB 1169, St. Louis, MO 63130, guinness@wustl.edu.

Introduction: The Planetary Data System (PDS) is in the midst of a multi-year effort to migrate its legacy datasets from the older PDS3 standard to the current PDS4 standard [1]. PDS4 is based on modern information modeling. It includes a rigorously defined information model for capturing metadata about planetary observations and has simpler and fewer data formats. The information model is implemented in XML. The PDS4 labels are strictly controlled by use of XML schema and Schematron such that labels are uniform in structure and content. The use of XML enables a variety of standard software tools to parse a PDS4 label and access the metadata about the data product.

Several currently active missions that archive data at the Geosciences Node (e.g., InSight, and Mars2020) already are using the PDS4 standard and all future mission data archives (e.g., Lunar Trailblazer and Europa Clipper) will use the standard. In addition, active missions will be converting their processing pipelines and delivering data in the PDS4 standard starting with new data collected in FY23. Thus, migrating legacy data will make the older data compatible with new planetary data. Another advantage of migration is that data discovery and processing tools only need to support a single archive standard.

The PDS Geosciences Node, led by Washington University, is migrating all its existing pre-PDS4 datasets as part of PDS-wide migration effort. This report will document the current status of Geosciences Node migration efforts and its plans for completing the task. Table 1 is a summary of the Node's migration efforts to date and plans going forward. The table shows the missions involved, the number of datasets, and estimated completion date.

Approach to Dataset Migration: We have developed two methods for migration to PDS4 depending on whether the existing data formats are compatible with the PDS4 standard. As noted PDS4 strives to allow simpler and fewer data product formats in order to streamline software needed to access and process the data. The PDS4 policy on allowable data formats can be found at: pds.nasa.gov/datastandards/documents/policy/format_policies_final.pdf.

If the data format for a dataset is compatible with the new PDS4 standard, then PDS4 labels are designed and generated for the existing data products. These new labels are integrated into the existing physical archive organization of the dataset. That is, data products will typically have parallel sets of PDS4 and PDS3 labels. The overlay approach will allow users to use their existing tools during the transition period. In addition to new data product labels, the required PDS4 bundle and collections products are added to the archive. New documentation also may be added to explain how the migration was done and to add any notes about issues found during migration. For example, the migration of MESSENGER data found a few errors in data files because PDS4 validation tools are more robust. These errors are listed in a document named 'geo_notes_and_errata' that was included in the migrated MESSENGER archives.

For the case where the data format is not compatible with PDS4 standards, the data products must be reformatted to conform to the PDS4 policy. PDS4 labels are designed and generated for the new data format, along with the required bundle and collection products, and any updated documentation. When migration is complete, the reformatted PDS4 archive will be peer reviewed as if it is a new dataset to make sure the new archive is scientifically useful and accurately captures the original data. An example of data that will need to be reformatted for the migration are Magellan Venus archives, which originally used binary encoding formats that are no longer in use.

Regardless of which migration approach is needed, all metadata about the data products will be included in the new PDS4 labels. In some cases additional metadata will be added to increase the findability and usability of the data. We plan on incorporating the migrated PDS4 version of each dataset into the Geosciences Node Orbital Data Explorer (ODE) and Analyst's Notebook (AN) data search and retrieval services once migration is complete.

Current Migration Status: The Geosciences Node migration work started with the MESSENGER and Mars Exploration Rover (MER) missions, which ended operations around the time that PDS4 migration efforts began. We started with these missions in order to capture any knowledge needed for migration before possibly important information was lost as project and instrument team personnel moved on to other projects. These mission migrations are complete.

Current migration efforts have focused on Lunar archives given the renewed international exploration of the Moon, in particular the CLPS and Artemis return to the Moon programs. As such, the Geosciences Node has completed the migration of its Chandrayaan-1 MiniRF, Lunar Prospector, LCROSS, and the higher-order GRAIL data. Note that the GRAIL raw science datasets will be migrated by the new PDS Radio Science Subnode. The Geosciences Node is currently working on migration of its Clementine archives (gravity and LIDAR datasets, and a derived brightness temperature dataset). These migrations should be complete by the end of FY23. The Node is also working with several instrument teams from the Lunar Reconnaissance Orbiter (LRO) mission to migrate their existing archives to PDS4 and to convert their data processing pipelines to generate PDS4 archives for new data acquired during their fifth extended mission. These include the LEND, LOLA, Diviner, and MiniRF archives. The migration of their existing archives should be completed by the end of FY23.

Future Migration Work: We will next focus on migration of Magellan Venus archives that were originally produced in the 1990s. Several Magellan datasets (e.g., MIDR, ARCDR, GxDR) are nearly ready for a peer review. These migrated datasets need to be peer reviewed because their data formats have changed. In the case of the MIDR and GxDR datasets, the small image frames that were designed to accommodate display systems available in the 1990s will be combined into larger mosaics. For the ARCDR dataset, the format is being updated because the data were originally stored as binary files using VAX floating point formats. The F-BIDR dataset will also be reformatted to a simpler format. We plan to provide a tool to allow users to make custom mosaics from multiple overlapping F-BIDR products.

As noted, active missions that the Geosciences Node is working - Mars Odyssey, Mars Science Laboratory (MSL), and Mars Reconnaissance (MRO) - will also be converting their processing pipelines to generate PDS4 for new data collected during their current extended missions. Most of the instrument teams on these missions will also be migrating their existing data to PDS4. The schedule for migration of existing data will depend on available team resources and the size of their archives. We have started working with several teams on this effort. For example, the MSL APXS team has already migrated their existing archive to PDS4 and is currently delivering new data in PDS4.

Finally, the schedule for migration of the remaining legacy non-PDS4 data is shown in Table 1. Most of these datasets have formats that do not require reformatting. There are a few cases where the format will need to be revised, notably the Mars Global Surveyor (MGS) MOLA altimetry profile dataset and the TES time sequential dataset. We expect to finish mission migrations by end of FY27 and datasets from individual data providers generated from data analysis programs by FY29.

References [1] Planetary Data System Standards Reference, Version 1.19.0.0, Oct. 1, 2022.

Table 1: PDS3 to PDS4 Migration Plan

Mission	Number of datasets	Estimated Completion
Past Missions		
MER	38	Complete
MESSENGER	14	Complete
Chandrayaan-1	4	Complete
Lunar Prospector	7	Complete
LCROSS	29	Complete
GRAIL	1	Complete
Clementine	6	In progress; FY23
Apollo	12	FY25
Magellan	15	In progress; FY24
Mars Phoenix Lander	12	FY25
Mars Global Surveyor	7	FY26
Mars Viking Lander	5	FY24
Mars Viking Orbiter	1	FY27
Mars Pathfinder	1	FY26
Active Missions		
LRO	18	In progress; FY23
MRO	17	In progress; FY25
MSL	18	In progress; FY24
Mars Odyssey	9	Best effort