INITIAL SUPPORT FOR PDS4 WITHIN THE GEOSPATIAL DATA ABSTRACTION LIBRARY (GDAL).
T.M. Hare, L.R. Gaddis and D.P. Mayer, U. S. Geological Survey, Astrogeology Science Center, Flagstaff, AZ, thare@usgs.gov.

Introduction: The NASA Planetary Data System (PDS) has released version 4 of their required archival format [1], now called PDS4. Since 2011, compliance with the PDS4 archiving standard has been required for data archives from NASA-funded missions and smaller research investigations. In contrast to previous versions that used the Object Description Language (ODL) format, PDS4 has been completely re-engineered. The most noticeable change for users will be the requirement for additional metadata and the switch to the eXtensible Markup Language (XML) format. A major goal for this move to PDS4 is to improve data discoverability by strongly enforcing consistency in the metadata and allowing for explicit cross-referencing for data products. Here we introduce initial support for PDS4 read/write capabilities within the Geospatial Data Abstraction Library (GDAL). Although this is only one of several options for preparing a PDS4-compliant archive, we believe that it is a valuable one for those who work with image and geospatial data.

GDAL: GDAL, released by the Open Source Geospatial Foundation (OSGeo), offers powerful capabilities for converting and processing geospatial planetary data. GDAL is a format translation library for geospatial raster and vector data [2]. In addition to the newly introduced PDS4 format, GDAL also supports PDS3 (read-only), USGS Astrogeology's image processing formats for the Integrated Software for Imagers and Spectrometers (ISIS2, ISIS3; read/write), Video Image Communication and Retrieval (VICAR; read) format [3], and the Flexible Image Transport System (FITS; read/write) format [4]. Mapping applications, which use the GDAL library for raster I/O (Input/Output), can also directly access these formats. Because GDAL supports more than 100+ formats, it is widely used across the terrestrial and planetary science community to easily manipulate and share geospatial data. For applications that do not use GDAL for I/O, the bundled routines released with GDAL can be used to convert common data formats into more universal geospatial formats (e.g., GeoTIFF).

PDS4 in GDAL Highlights: The PDS4 driver in GDAL currently supports simple reading and writing of PDS4 raster-based (image) products (see Section 4.2.1 in the PDS4 Data Providers Handbook https://pds.nasa.gov/pds4/doc/dph/current/). Highlights of the GDAL PDS4 driver are listed below.

- **Release Date**: Initial PDS4 support in GDAL was released for testing in September of 2017 (http://www.gdal.org/frmt_pds4.html), when Hobu Inc. completed the first of two planned contracts.
- **Templates for PDS4 Creation**: Currently GDAL only automates translation of basic properties of the input image (lines, samples, bit type, etc.) and map projection parameters. The PDS4 GDAL driver alone lacks the ability to create or propagate much of the required metadata for writing a PDS4 compliant data product. Thus, a complete solution using GDAL will require a well-developed PDS4 XML templates or the creation of simple scripts to retrieve specific values (e.g., from existing file labels such as those from PDS3 or ISIS3 images), to calculate, translate and otherwise add additional metadata. These PDS4 XML templates must meet PDS4 formatting standards and organization and provide information not propagated by GDAL during format conversion (e.g., author, institute name, processing details). Further, use of PDS4 XML templates must be tailored for each mission and/or image product type. At the current time, the collection of such label templates for a wide variety of image products is limited but new or updated examples are being regularly added (see: https://pds.nasa.gov/pds4/about/portal.shtml).

  To help support more automated creation of a compliant PDS4 label during a format conversion, GDAL supports user-defined template variables. This allows the user to update these defined variables (e.g., start time, mission phase, etc.) via the command-line or within a simple script.
- **Remote PDS4 Templates**: To facilitate PDS4 template sharing, GDAL supports loading remote templates from an http address. Thus, PDS nodes, mission teams, and researchers will be able to host example templates from their own website or from sites like Github.
- **Low-level GDAL API**: Full XML access during a format conversion is made available using C++ (Xerces library) or a Python (lxml library) application protocol interface (API). This low-level access to the PDS4 label supports development of more robust software applications other than simple scripts.
- **Interoperability**: Because of the PDS4 requirement for extensive metadata, it is difficult for the PDS4 format to be interoperable across many applications. To facilitate PDS4 interoperability, GDAL supports writing a detached PDS4 label that references a raw pixel stream from an uncompressed planetary GeoTiff (other formats like FITS or VICAR could also be

easily added). Because GeoTiff has broad support across many mapping and scientific applications, this allows for development of a PDS4-compliant archive and supports use of a more interoperable and universally used scientific image file format at the same time.

- **Direct Mapping Support**: By supporting format drivers in GDAL, the PDS4 reader will eventually find its way into GIS applications like QGIS, UDIG GIS, and Saga GIS. There are plans to also add read support for Esri's ArcGIS Pro and possibly ArcMap.

PDS4 in GDAL Limitations: There are several limitations to the current PDS4 support within GDAL.

- GDAL is primarily a geospatial library and thus the current driver is focused on supporting high-level derived (map projected) data sets. Low-level (engineering or EDRs) data can be supported by using PDS4 XML templates and other PDS tools.
- Because GDAL automates propagation of the basic properties of the input image and map projection parameters, if one needs to add more metadata into those sections (e.g., the band's unit designation), a user will need to manually update the label or run a post-processing script. We plan to include options to support this within the planned second contract.
- There is currently no support for PDS4 table files. This will also be added during the next contract. Like the GDAL PDS3 driver, if "Latitude" and "Longitude" fields are defined, the table will be treated as a geospatial vector point layer, which is suitable for direct display or conversion in many mapping applications.
- There is currently no support for writing multiple image arrays into one file (called sub-datasets within GDAL). This should not be confused with writing multiple-band images, which is supported. This capability also will be added during the second contract by using an "append" to an existing PDS4 file. This means that support for PDS4 "composite" headers will need to be handled within the initial master PDS4 template.

Summary: The currently envisioned GDAL workflow for supporting PDS4 image format conversions will rely heavily on user-tailored PDS4 XML label templates and/or user scripts or applications. GDAL, first-and-foremost, is a library to write code against, whether it is for reading, writing or translating across the 100+ supported image formats. Thus, an archiving workflow for PDS4 will require up-front archive layout, label design, and metadata input from the data provider. When designing the PDS4 driver for GDAL, the challenge was to create a useful tool, which, most importantly, allows for direct read support (for application I/O and conversion), while also supporting a flexible PDS4 creation solution when partnered with PDS4 XML templates via user scripts and applications. This tool is being actively used by the PDS Cartography and Imaging Sciences node, and sample labels and processing scripts will soon be available.

Challenges for PDS4: Understanding and using the PDS4 standard and information model remains demanding for users. To keep the format straightforward to maintain within a PDS archive designed for long-term preservation, only four basic structural data formats are allowed (including 2-D arrays with binary data, tables as repeating records, parsable byte streams and encoded byte streams). Compressed file formats (e.g., Jpeg2000, used by NASA's MRO HiRISE) are no longer allowed. Not only will a PDS4 migrated Jpeg2000 file be significantly larger, beneficial aspects like built-in pyramids (for quick rendering) and streaming capabilities are also not currently supported by PDS4. Because of significant issues like these, PDS3-formatted data will remain available in PDS archives for years into the future.

Support for complicated vector file formats, as those used in GIS mapping applications, are also challenging to support in PDS4. Currently we are researching an XML vector format, the Geographic Markup Language (GML), which is both ASCII-based and yet robust enough to support GIS vectors (points, lines, and polygons; see http://bit.ly/2ALQDf0).

Future Updates: While the PDS4 standard has been available for several years, it is still being updated as issues are encountered and capabilities added. We have planned for a future GDAL contract in 2018 or early 2019 to help keep up with these changes and to include capabilities not yet completed. This includes sub-dataset, table and perhaps GML vector-based revisions as discussed above.

Acknowledgments: This effort was supported by NASA's Planetary Spatial Data Infrastructure (PSDI) InterAgency Agreement and the PDS Cartography and Imaging Sciences Node.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References: [1] PDS4 Information at the Planetary Data System, URL: https://pds.nasa.gov/pds4/. [2] Hare, T.M., et al., 2007, LPSC 39, abs #2536. [3] Gaddis, L.R., Hare, T., and Beyer, R., 2014, Summary and Abstracts of the Planetary Data Workshop, June 2012, U.S. Geological Survey Open-File Report 2014–1056, page 199. [4] Marmo, C. et al., 2018, Planetary Science Informatics and Data Analytics, St. Louis, MO.