

Notes and Errata Regarding MESSENGER PDS3 to PDS4 Migration

PDS Geosciences Node

October 18, 2018

Notes

These notes apply to MESSENGER archives at the PDS Geosciences Node, specifically the GRS, NS, MLA, XRS, and derived Radio Science data. All MESSENGER data sets were originally archived under the PDS3 standard. The PDS Nodes hosting MESSENGER archives have migrated the data to the PDS4 standard.

With a few exceptions, the data products in the PDS3 MESSENGER archives at the Geosciences Node are already PDS4-compliant, so there was no need to alter the data files. The PDS3 archive volumes remain intact; no files have been removed. The only change is the addition of PDS4 labels and documentation files. (See also **Data Products That Are Not PDS4-Compliant** below.) In the data directories, the metadata in PDS3 labels has been copied to PDS4 labels, so that each data product now has both a PDS3 and a PDS4 label.

Instead of the data sets and archive volumes in PDS3, products in PDS4 are organized into collections and bundles. A collection is a set of related products, which may be data products, document products, browse products, miscellaneous products, etc. A bundle is a set of related collections.

For these MESSENGER archives a PDS4 bundle has been defined for each PDS3 volume. Each bundle has one or more data collections, a document collection, and a miscellaneous collection that holds the PDS3 index table. Some bundles also have a browse collection that holds browse products in the extras directory. The table below shows the correspondence between PDS3 volumes and PDS4 bundles.

PDS4 Bundles and Corresponding PDS3 Volumes and Data Sets

Archive	PDS4 Bundle	PDS3 Volume	PDS3 Data Set ID
GRS Raw Data	urn:nasa:pds:mess_grs_raw	messgrs_1001	MESS-E/V/H-GRNS-2-GRS-RAWDATA-V1.0
GRS Calibrated and Derived Data	urn:nasa:pds:mess_grs_derived	messgrs_2001	MESS-E/V/H-GRNS-3-GRS-CDR-V1.0
NS Raw Data	urn:nasa:pds:mess_ns_raw	messns_1001	MESS-E/V/H-GRNS-2-NS-RAWDATA-V1.0
NS Calibrated and Derived Data	urn:nasa:pds:mess_ns_derived	messns_2101	MESS-E/V/H-GRNS-3-NS-CDR-V2.0, MESS-E/V/H-GRNS-4-NS-DDR-V2.0
MLA Raw Data	urn:nasa:pds:mess_mla_raw	messmla_1001	MESS-E/V/H-MLA-2-EDR-RAWDATA-V1.0
MLA Calibrated, Derived, and Map Data, V1.0	urn:nasa:pds:mess_mla_calibrated	messmla_2001	MESS-E/V/H-MLA-3/4-CDR/RDR-DATA-V1.0, MESS-E/V/H-MLA-4-GDR-V1.0
MLA Calibrated, Derived, and Map Data, V2.0	urn:nasa:pds:mess_mla_derived	messmla_2101	MESS-E/V/H-MLA-3/4-CDR/RDR-DATA-V2.0, MESS-E/V/H-MLA-4-GDR-V2.0
XRS Raw Data	urn:nasa:pds:mess_xrs_raw	messxrs_1001	MESS-E/V/H-XRS-2-EDR-RAWDATA-V1.0

Archive	PDS4 Bundle	PDS3 Volume	PDS3 Data Set ID
XRS Calibrated Data	urn:nasa:pds:mess_xrs_calibrated	messxrs_2001	MESS-E/V/H-XRS-3-CDR-SPECTRA-V1.0
XRS Maps and Footprint Data	urn:nasa:pds:mess_xrs_derived	messxrs_3001	MESS-H-XRS-3-RDR-MAPS-V1.0, MESS-H-XRS-5-RDR-FOOTPRINTS-V1.0
Radio Science Derived Data (OCC, SHADR, SHBDR)	urn:nasa:pds:mess_rs_derived	messrs_1001	MESS-H-RSS/MLA-5-SDP-V1.0

A bundle is identified by a file named `bundle_*.xml` in the root directory of a volume; it describes the bundle and lists the collections that belong to it. A collection is identified by a file named `collection_*.xml` in a subdirectory. The file `collection_*.inventory.csv` is a list of the products that belong to the collection.

Every product, collection, and bundle in PDS4 has a Logical Identifier (LID) which is guaranteed to be unique throughout PDS. The LID is defined in the PDS4 label using the `<logical_identifier>` tag. For data products, the LID is analogous to `PRODUCT_ID` in a PDS3 label.

Not every PDS3 directory has a PDS4 counterpart. The PDS3 data, document, and index directories have been made into collections, but the catalog, geometry, and label directories have been left unchanged. In order to preserve the information in the PDS3 catalog files, the text has been copied to `*.txt` files in the document directories and given PDS4 labels. The files **aareadme.txt** and **errata.txt** in the volume root directories have been left alone, except that a brief note has been added to the beginning of **aareadme.txt** directing the reader to the PDS4 file **readme.txt** in the bundle root directory. No PDS4 labels have been provided for **aareadme.txt** and **errata.txt**.

Data Products That Are Not PDS4-Compliant

The only data products in these PDS3 archives that are not PDS4-compliant are JPEG-2000 images (`*.JP2`). These images have been converted to standard PDS images (`*.IMG`). Their PDS4 labels refer to the IMG file as the primary data product and the JP2 file as a supplemental product.

The conversion from JP2 to IMG was done using GDAL (<https://www.gdal.org>) using a command like this:

```
gdal_translate -of ENVI grs_dap_k_abd_map.jp2 grs_dap_k_abd_map.img
```

The `-of ENVI` switch produces a PDS4-compliant output file with no compression and no metadata. It does produce an ENVI header file, which is not included in the archive.

For verification, we loaded the JP2 and IMG products into ENVI, subtracted one from the other, and confirmed that the resulting raster contained only zeroes.

This conversion was done for 23 images in the GRS and XRS archives. The MLA archives also contain JP2 files, but they already have IMG counterparts, so no conversion was needed.

Errata

The following errors and anomalies were discovered during the migration process. If errors were found in the original PDS3 labels, they were corrected in the PDS4 labels where possible. If errors were found in the data files, they were left unchanged.

1. **Mission Phase Name.** The MLA map products were generated from data acquired during multiple mission phases, which are listed in the PDS3 labels. At the time of this migration, the PDS4 mission dictionary for MESSENGER allowed only one mission phase name to appear in a PDS4 label. Therefore the mission_phase_name attribute was omitted from the PDS4 labels for these products.
2. **Instrument Mode ID.** The PDS3 labels for the MLA derived data products included the keyword INSTRUMENT_MODE_ID, which has no counterpart in the PDS4 core dictionary or the MESSENGER mission dictionary. This keyword and value are not present in the PDS4 labels.
3. **Detector ID.** The GRS High Purity Germanium Detector has the abbreviation HPGe, using mixed case. The MESSENGER mission dictionary specifies this value erroneously as “hpge”, using lower case. Therefore this value appears in the PDS4 labels in lower case.
4. **ASCII real number representation.** The PDS core dictionary at the time of this migration did not allow ASCII real number values ending in a decimal point, e.g. “120.”. This notation appears in many MESSENGER data products. A Standards Change Request to allow a real number ending in a decimal point was submitted to PDS. Until this change is made, the use of this notation will cause an error in validation software that compares PDS4 labels and ASCII tables to the current PDS core dictionary.
5. **White space between columns.** In NS CDR GAB and NS DDR fixed-width ASCII tables, white space appears between table columns. The PDS4 labels include fields named “null” to account for the space characters.
6. **Asterisks in ASCII table data.** A few data files were found to have fields containing a series of asterisk characters, e.g. “*****”, where they should have contained an ASCII real number. This error occurs only in a few cells in a few tables, not systematically. The affected files are in the NS derived and MLA calibrated data.
7. **Undocumented value for CHANID.** In MLA derived products mlasccdr1104081536.tab and mlasccdr1103301417.tab, the CHANID field has the value of “E”, which is not documented in the label or the SIS. This field was defined as ASCII_INTEGER in PDS3 labels, even though its description allowed values of A, B, and C in addition to integers 0 through 9. In PDS4 labels the field has been defined as ASCII_String, but it still lacks a definition for the E value.
8. **Wrong data type in PDS3 label.** In PDS3 labels for GRS DAP products, the SAMPLE_TYPE and SAMPLE_BITS for the uncompressed image are listed as IEEE_REAL and 32, respectively, but the uncompressed image was found to have pixels of unsigned bytes. The PDS4 labels correctly list the data_type as UnsignedByte.
9. **Wrong image dimensions in PDS3 label.** The XRS RDR Map products have incorrect dimensions given in their PDS3 labels. The UNCOMPRESSED_FILE sections of these labels list that the image's dimensions are 1440 lines by 720 samples. These values should be 720 lines by 1440 samples. The PDS4 labels have the correct dimensions.

10. **Wrong number of records in PDS3 labels for RS products.** The PDS3 labels for several of the RS SHBDR products listed the wrong number of records in the tables of the products. The correct values were determined and included in the PDS4 labels.
11. **Wrong number of records in PDS3 labels for MLA products.** In PDS3 labels for sixteen MLA STACDR products, the values for FILE_RECORDS and ROWS are wrong. The PDS4 labels list the correct number of records. The affected labels are as follows, all in the cdr/2011/dec and cdr/2012/jan/ directories.

mlastacdr1112311701.tab
mlastacdr1201010449.tab
mlastacdr1201011637.tab
mlastacdr1201020424.tab
mlastacdr1201021602.tab
mlastacdr1201030350.tab
mlastacdr1201031538.tab
mlastacdr1201040326.tab
mlastacdr1201050302.tab
mlastacdr1201051450.tab
mlastacdr1201060237.tab
mlastacdr1201061425.tab
mlastacdr1201070212.tab
mlastacdr1201071400.tab
mlastacdr1201080148.tab
mlastacdr1201081335.tab

12. **Wrong values for Boolean data type.** In PDS3 labels for MLA HADCDR products, several fields were defined as type BOOLEAN with values "T" and "F". In PDS4, ASCII_Boolean values are required to be "True" and "False", so these fields were defined in the PDS4 labels as ASCII_String instead.
13. **UTC time missing final "Z".** In PDS3 labels for MLA RADR products, field 17 was defined as UTC time, but the values did not end in the "Z" character required for UTC times. So the PDS4 labels define this field as ASCII_Date_Time_YMD instead of ASCII_Date_Time_YMD_UTC.
14. **UTC time missing final "Z" and four digits after the decimal point.** In PDS3 labels for MLA SCIRDR products, several fields defined as having UTC times did not end in "Z", and also had 4 digits after the decimal point for the number of seconds. The PDS4 standard at the time of this migration was to have no more than 3 digits after the decimal point. So the PDS4 labels define these fields as ASCII_String.
15. **Wrong contents of data file.** The NS CDR data file data\cdr\2010\04\ns_cdr_gab2010100zzz.tab was found not to contain any data, but instead to contain a copy of the PDS3 label file. Our checksum records show that this is the way the file was delivered to us originally. As we did not have content validation tools for PDS3 deliveries, our standard validation procedure would not have found this error. We have asked the data providers to send the data file.