

Clementine Bistatic Radar Intermediate and Reduced Data: Notes and Errata Regarding PDS3-to-PDS4 Migration

PDS Geosciences Node

May 28, 2024

Notes

All Clementine data sets are archived as part of NASA's Planetary Data System (PDS). These data sets were originally archived under the PDS3 standard, and the PDS Nodes hosting Clementine archives are actively migrating these archives to the PDS4 standard.

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.

Hereinafter, these notes apply specifically to the Clementine archive at the PDS Geosciences Node that contains intermediate and reduced bistatic radar data. The original PDS3 archive was split into eight volumes, sized so that each would fit on a CD-ROM. These eight volumes are combined into a single PDS4 bundle with multiple data collections and document collections. The PDS4 bundle is stored in a separate top-level directory from the original PDS3 volumes. The PDS3 volumes are still available in the Geosciences Node site at the time of the migration. The table below shows the correspondence between each PDS3 volume and the PDS4 bundle.

PDS4 Bundles and Corresponding PDS3 Volumes and Data Sets

Archive	PDS4 Bundle	PDS3 Volumes	PDS3 Data Set ID
Clementine Bistatic Radar Intermediate and Reduced Data	urn:nasa:pds:clementine1_bistatic	CL_2101, CL_2102, CL_2103, CL_2104, CL_2105, CL_2106, CL_2107, CL_2108	CLEM1-L-RSS-5-BSR-V1.0

Across the original volumes, some files were duplicated, accumulated, or revised. Thus, in combining these volumes into a single PDS4 bundle, the following actions were taken:

1. Directories from each volume are combined at the root level of the bundle.
 - a. Example: The contents of each volume's /img directory now populate the bundle's single /img directory.
 - b. Note: As a single special case, both the /document and /extras directories from each volume are merged into the bundle's root-level /document directory.
2. For improved navigation, the bundle's /img directory is organized by observation time, or equivalently, by the first five characters of each file name.
 - a. Example: cl_2101/img/66396val.img was migrated to img/66396/66396val.img.
3. For each file exactly duplicated across all volumes, only one representative is included in the bundle.
4. Four files were exactly duplicated across all volumes except in the final (latest) volume, where they were updated. Only the final (updated) version of each of these files is retained in the bundle.

- a. Files: catalog/dataset.cat, document/docinfo.txt, document/lpsc_abs.lbl, and document/procssng.txt.
 - b. Other than dates, the only differences are typographical corrections, as described in document/errata.txt.
5. Each volume contains both a per-volume and cumulative index. In the bundle, only the final, complete cumulative index is retained: index/cumindex.tab.
 - a. Note that this cumulative index includes the original volume for each file, so that each per-volume and progressively cumulative index can be parsed from it.
 - b. Note also that we treat cumindex.tab as a historical record of the original archive organization, which may be useful in interpreting path references in the documentation. It has not been updated to reflect the reorganization performed during migration.
6. Two root-level files are nearly identical across volumes. Only the final version of each of these files is retained in the bundle.
 - a. Files: aareadme.txt, voldesc.cat
 - b. The only differences across volumes are volume references (in the title and/or PDS3 label) and publication dates (which vary by up to one day).
7. The file errata.txt is only populated in the final volume, and therefore only that version is retained in the bundle. Additionally, whereas errata.txt is in the root level of each volume, the file resides at document/errata.txt in the bundle.

The only additions in the bundle relative to the PDS3 volumes are

1. a brief note at the beginning of aareadme.txt directing the reader to the PDS4 readme file (readme_clementine1_bistatic.txt) in the bundle's root directory
2. PDS4 labels (metadata in PDS3 labels have been copied to PDS4 labels, so that each migrated data product now has both a PDS3 and a PDS4 label)
3. PDS4 documentation files
4. the file sort/srtnpwr_reconstructed.tab (see Errata section below)

The volumes also contain two PDF files. The original files are not PDS4 compliant, but each has been converted to a compliant PDF and may be found at document/lpsc_abs.pdf and document/offsqnt.pdf. In addition, the file extension of sort/lcp.tab and sort/rcp.tab is .img in the original PDS3 archive. In that archive, these files are described as ASCII "images" for which each "pixel" is a numeric value written as a string. However, PDS4 does not support such ASCII images, and therefore these files are now instead archived as tables. The text of the PDS4 label for each of these files is slightly edited to reflect this change, and the file extension of the files themselves is changed to .tab to avoid confusion.

Except as described in the foregoing paragraphs, the contents of the PDS3 archive are intact.

A bundle is identified by a file named bundle_*.xml (where * is some character sequence) in the root directory of a bundle. This PDS4 label describes the bundle and lists the collections that belong to it. A collection, in turn, is identified by a PDS4 label named collection_*.xml in a subdirectory. The file collection_*_inventory.csv is found in the same directory as the collection label and lists the products that belong to that collection. Each product also has its own PDS4 label named *.xml.

Every bundle, collection, and product in PDS4 has a Logical Identifier (LID) that is guaranteed to be unique throughout the 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 component has a PDS4 counterpart. Namely, PDS3 readme (aareadme.txt), volume description (voldesc.cat), map (catalog/dsmmap.cat), and person (catalog/person.cat) files do not have PDS4 labels.

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. **Parsing corrections.** For several files, the parsing details in the PDS3 labels are incorrect and have been corrected in the PDS4 labels. Specifically:
 - a. data/*.tab (8 files)
 - i. The start of the COMMENT column is offset by 16 bytes relative to the start stated in the PDS3 label. If the start given in the PDS3 label were used instead, this ASCII field would contain some non-printable ASCII characters (unexpected) and some non-ASCII characters (invalid). The true location of the field was determined by isolating all printable ASCII characters in the record that come after the penultimate field, as COMMENT is the last field in the (single-record) header table. All such characters are found in a contiguous interval 80 characters in length, which is the length stated in the PDS3 label. The field contents, thusly parsed, were also inspected and confirmed to be human-readable and meaningful.
 - ii. The bulk of each file is the data table, which has 128 columns that each contain 187,500 data samples. The PDS3 label describes these columns as being 128 bytes wide and their data type as real. However, the PDS4 label instead describes these columns as 128 *bits* wide and *complex*. The columns, thusly parsed, exactly fit the size allocated to the table in the file. Moreover, the file, table, and field descriptions in the PDS3 label (which are preserved in the PDS4 label) all describe the data as complex. Finally, the values, thusly parsed, were inspected and their magnitudes confirmed to be reasonable (that is, neither extremely large nor extremely small; extreme magnitudes would be expected for incorrectly parsed binary numeric data).
 - b. index/cumindex.tab
 - i. Both column starts and widths are incorrect in the PDS3 label. As this file is ASCII and columns are comma-delimited, determining the correct starts and widths was straightforward and unambiguous.
 - c. sort/lcp.img, sort/rcp.img
 - i. These two files describe their contained values as being 16 bits each. However, as each file is ASCII and the values are clearly delimited by spaces, it is straightforward and unambiguous to identify the correct length as 16 *bytes*. Moreover, thusly parsed, the dimensions of each table are 1024x1464, as stated in the PDS3 label.
 - d. sort/srtnpwr.tab
 - i. The PDS3 label states that there are 72 columns in the NUMBER OF VALID POINTS field, but only 63 are present. As this file is ASCII and columns are comma-delimited, determining the correct column count was straightforward and unambiguous.

2. **Incomplete data.** As stated above, sort/srtnpwr.tab has 63 columns in its NUMBER OF VALID POINTS field even though its PDS3 label states that the field has 72 columns. This difference is not merely a typographical error. Rather, each column corresponds to a target, and nine columns (targets) are missing. We have plausibly reconstructed the complete 72-column table as sort/srtnpwr_reconstructed.tab. Details of this reconstruction are provided in the PDS4 label sort/srtnpwr_reconstructed.xml.