Notes on the Migration of Mars Science Laboratory (MSL) CheMin (CHMN) Raw Data from PDS3 to PDS4

PDS Geosciences Node

November 6, 2023

These notes apply to the transition of Mars Science Laboratory (MSL) CheMin (CHMN) raw data from PDS3- to PDS4-compliant data deliveries beginning with MSL PDS data delivery #34, and the migration of earlier CHMN PDS data from PDS3 to PDS4.

Notes on New Releases

Beginning with MSL PDS Release 34, December 5, 2023, new CHMN raw data products are archived under the PDS4 standard. The older data are planned to be migrated at a later date, but for now the older data directories remain unaltered (see "Notes on Migrated Data" below).

No files have been removed from the PDS3 archive volume. The only change is the addition of PDS4 labels and documentation files. New CHMN data products include both PDS3 and PDS4 labels. Since the CHMN data products are already PDS4-compliant, no change has been made to the data product format.

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.

One PDS4 bundle has been defined for the one dataset on the CheMin PDS3 EDR volume. The CheMin raw bundle has one data collection, two document collections, and two miscellaneous collections. The tables below show the correspondence between the PDS3 volume and PDS4 bundle and PDS3 product types and PDS4 collection, respectively.

MSL CheMin Raw PDS4 Bundle and Corresponding PDS3 Volume and Data Sets

Archive	PDS4 Bundle	PDS3 Volume	PDS3 Data Set ID
MSL CheMin Raw Data	urn:nasa:pds:msl_chemin_raw	mslcmn_0xxx	MSL-M-CHEMIN-2-EDR-V1.0

MSL CheMin Raw PDS4 Collection and Corresponding PDS3 Product Types

PDS4 Collection	PDS4 Collection ID	PDS3 Product Type
Raw	data	CHEMIN_ECC CHEMIN_ED1 CHEMIN_EDA CHEMIN_EDS CHEMIN_EE1 CHEMIN_EEA CHEMIN_EES CHEMIN_EHK CHEMIN_ETR

A bundle is identified by a file named **bundle_*.xml** in the root directory of a bundle; 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 tag **<logical_identifier>**. 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, catalog, document, index, and software directories have been made into collections, but the calib and label directories have been left unchanged. No PDS4 labels have been provided for **aareadme.txt, voldesc.cat,** and **errata.txt** in the volume root directories. Errata.txt has been replaced by a release notes file (in the document collection), which will be updated with each release. Note that the SIS documents in the document directory refer to the PDS3 datasets.

Notes on Migrated Data

Data products in PDS3 CheMin archives at the PDS Geosciences Node are already PDS4-compliant, so there will be no need to alter the data format when the older files are migrated. In the data directories, the metadata in PDS3 labels will be copied to PDS4 labels, so that each data product will have both a PDS3 and a PDS4 label. At the time of writing the completion date for this migration has not been scheduled but is anticipated in 2024.

Exclusion of CheMin FILM data products from PDS4 deliveries

CheMin FILM (EFM) products have been submitted in PDS3 data deliveries but the FILM format is incompatible with PDS4 (data in a FILM readout of the CheMin CCD are 20-bit unsigned integers, a format not accommodated in PDS4). A FILM product is a crude, brute-force use of the CCD without using the energy selection capability that is crucial to collection quality XRD data; the only use for the FILM product would be as a last-ditch attempt to extract XRD data if energy selection by the CCD were to fail completely. The assumption before TVAC and surface ops was that some of the strongest Debye diffraction rings might be discernable above the very high total-energy fluorescence background. In practice the FILM image is far too noisy and its use is moot. Exclusion of FILM data is of no impact to in-practice CheMin operation, but these files will continue to be stored outside PDS in the Astrobiology Habitable Environments Database (http://odr.io/CheMin).