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

PDS Geosciences Node

June 7, 2023

These notes apply to the transition of Mars Science Laboratory (MSL) CheMin (CHMN) data from PDS3- to PDS4-compliant data deliveries beginning with MSL PDS data delivery #33, and the migration of earlier CHMN PDS data from PDS3 to PDS4.

Notes on New Releases

Beginning with MSL PDS Release 33, August 1, 2023, new CHMN 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 will be addition of PDS4 labels and documentation files. New CHMN data products will include both PDS3 and PDS4 labels.

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 two datasets on the CheMin PDS3 RDR volume. The CheMin derived bundle has three data collections, 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 collections, respectively.

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

Archive	PDS4 Bundle	PDS3 Volume	PDS3 Data Set ID
MSL CheMin Derived Data	urn:nasa:pds: msl_chemin_derived	mslcmn_1xxx	MSL-M-CHEMIN-4-RDR-V1.0, MSL-M-CHEMIN-5-RDR-V1.0

MSL CheMin PDS4 Collections and Corresponding PDS3 Product Types

PDS4 Collection	PDS4 Collection ID	PDS3 Product Type
RE1 (single-pixel XRF energy)	data_chemin_re1	CHEMIN_RE1
RDA (XRD diffraction-all)	data_chemin_rda	CHEMIN_RDA
MIN (qualitive mineralogy)	data_chemin_min	CHEMIN_MIN

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, extras, and index 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.

The three data collections correspond to the three Reduced Data Records (RDRs) that meet the science goals of the CheMin instrument. The RDRs are processed from corresponding EDR sources as shown in the table below:

RDR	Data source	Description
RE1 (single-pixel XRF energy)	EE1 EDRs	XRF energy summed from those single pixels surrounded by pixels below an energy threshold. Processed by converting raw DN energy to keV (reversible processing).
RDA (XRD diffraction-all)	EDA EDRs	1D XRD plots of 2-theta versus intensity obtained by integration of the 2D Debye diffraction rings radially from the beam stop. Processing is not reversible (the original 2D pattern cannot be backed out of the 1D pattern).
MIN (quantitative mineralogy)	Higher-level analysis of an RDA (CODMAC level 5; NASA level2)	Mineralogy determined by use of commercial or freely available software packages (e.g., JADE, FULLPAT) to analyze an RDA file (non-reversible processing).
Note: The RE1 files provide highest quality XRF data and the RDA files provide highest quality XRD data. These two formats are used exclusively to prepare CheMin reduced XRF and XRD files. For details of these and other formats refer to Section 2.3 of the CheMin RDR SIS.		

A typical CheMin PDS4 delivery will include the combined RE1, RDA, and MIN files from a given drill or scoop sample with sample data (plus "background" RE1 files of empty or emptied cells), along with the following three documents: (1) release notes, (2) an index table, and (3) a label file for the index table.

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 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.

PDS4 XML labels will be prepared for previously archived CheMin data from MSL data deliveries #1 to #32 (February 2013 to March 2023). At the time of writing the completion date for this migration has not been scheduled but is anticipated later in 2023, or in 2024. As noted above, existing CheMin RDR csv-format data files

for these earlier deliveries are already PDS4 compliant and will not be modified unless corrections of errata are required.

Exclusion of CheMin FILM data products from PDS4 deliveries

CheMin FILM 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).