

Mars 2020
SHERLOC
PDS Archive Bundle
Software Interface Specification (SIS)

Version 1.0

July 30, 2021

Prepared by: Susan Slavney

Custodian: Susan Slavney

Mars 2020
SHERLOC
PDS Archive Bundle
Software Interface Specification (SIS)

Version 1.0

July 30, 2021

Custodian:

Susan Slavney
PDS Geosciences Node, Washington University in St. Louis

<date>

Approved:

Amy Chen
Mars 2020 SHERLOC Archive Representative, JPL

<date>

Deborah Padgett
Mars 2020 Instrument Data Services, Jet Propulsion Laboratory

<date>

Table of Contents

Tables.....	iv
Figures.....	iv
Document Change Log.....	iv
TBD Items.....	iv
Acronyms and Abbreviations.....	v
Glossary.....	vi
1 Overview.....	1
1.1 Purpose and Scope.....	1
1.2 Contents.....	1
1.3 Applicable Documents.....	1
1.4 Audience.....	2
1.5 Mars 2020 Mission.....	2
1.6 SHERLOC Description.....	2
2 SHERLOC Data Products.....	1
2.1 Data Product Overview.....	1
2.2 Data Processing Levels.....	1
3 SHERLOC Archive Organization.....	1
3.1 The SHERLOC Bundle.....	1
3.2 SHERLOC Collections.....	3
3.3 SHERLOC Data Organization.....	3
3.4 SHERLOC Product Identification and Naming.....	3
3.4.1 Logical Identifiers.....	4
3.4.2 File Naming Convention.....	5
3.5 PDS4 Labels.....	8
3.6 PDS4 Data Dictionaries.....	8
4 SHERLOC Product Formats.....	9
4.1 Data Product Formats.....	9
4.2 Document Product Formats.....	9
Appendix A Support Staff and Cognizant Persons.....	10
Appendix B Bundle Directory Structure.....	11

Tables

Table 1 All SHERLOC Data Product Types..... 1
 Table 2 Data Processing Level Definitions 1
 Table 3 Collections in the SHERLOC Bundle..... 3
 Table 4 SHERLOC File Name Components 5
 Table 5 PDS4 Dictionaries Used In SHERLOC Labels 9

Figures

Figure 1 PDS bundles, collections, and products. 2
 Figure 2 Mars 2020 File Naming Scheme..... 5

Document Change Log

Version	Change	Date	Affected portion
0.9	Pre-peer-review version	January 26, 2021	All
1.0	Post-peer-review version	April 29, 2021	Table numbering, Table 1, Table 4, 1.3, 3.4, Appx B, TBD Items
1.0	Updated for Release 1	July 30, 2021	1.3, Applicable Documents 3.4.1, Logical Identifiers Appx B

TBD Items

Item	Section(s)	Responsibility

Acronyms and Abbreviations

Acronym/Abbreviation	Meaning
ASCII	American Standard Code for Information Interchange
EDR	Experiment Data Record
FSW	Flight Software
GDS	Ground Data System
HTML	HyperText Markup Language
IDS	Instrument Data System
JPL	Jet Propulsion Laboratory
LID	Logical Identifier
LIDVID	Versioned Logical Identifier (logical identifier with version identifier)
N/A	Not Applicable
NASA	National Aeronautics and Space Administration
NSSDCA	National Space Science Data Coordinated Archive
PDS	Planetary Data System (the organization)
PDS4	Planetary Data System Version 4 (the archive standard)
RCE	Rover Compute Element
RDR	Reduced Data Record
SCLK	Spacecraft Clock
SFDU	Standard Format Data Unit
SHERLOC	Scanning Habitable Environments with Raman & Luminescence for Organics & Chemicals
SIS	Software Interface Specification
Sol	Mars solar day
SPICE	Spacecraft, Planet, Instrument, C-matrix, Events kernels
TBD	To Be Determined/Defined
TBPB	To Be Provided By
UTC	Coordinated Universal Time
VICAR	Video Image Communication and Retrieval

VID	Version Identifier
XML	Extensible Markup Language

Glossary

Many of these definitions are taken from Appendix A of the PDS4 (Planetary Data System Version 4) Concepts Document, pds.nasa.gov/pds4/doc/concepts. The reader is referred to that document for more information.

Archive – A place in which public records or historical documents are preserved; also the material preserved, often used in plural. The term may be capitalized when referring to all of PDS holdings (i.e., the PDS Archive).

Basic Product – The simplest product in PDS4; one or more data objects (and their description objects), which constitute (typically) a single observation, document, etc. The only PDS4 products that are *not* basic products are collection and bundle products.

Bundle – A list of related collections. For example, a bundle could list a collection of raw data obtained by an instrument during its mission lifetime, a collection of the calibration products associated with the instrument, and a collection of all documentation relevant to the first two collections.

Class – The set of attributes (including a name and identifier) which describes an item defined in the PDS Information Model. A class is generic, i.e., a template from which individual items may be constructed.

Collection – A list of closely related basic products of a single type (e.g. observational data, browse files, documents, etc.). A collection is itself a product (because it is simply a list, with its label), but it is not a *basic* product.

Data Object – A generic term for an object that is described by a description object. Data objects include both digital and non-digital objects.

Description Object – An object that describes another object. As appropriate, it will have structural and descriptive components. In PDS4 a ‘description object’ is a digital object, such as a string of bits with a predefined structure.

Digital Object – An object which consists of electronically stored (digital) data.

Identifier – A unique character string by which a product, object, or other entity may be identified and located. Identifiers can be global, in which case they are unique across all of PDS (and its federation partners). A local identifier must be unique within a label.

Label – The aggregation of one or more description objects such that the aggregation describes a single PDS product. In the PDS4 implementation, labels are constructed using XML (eXtensible Markup Language).

Logical Identifier (LID) – An identifier that identifies the set of all versions of a product.

Versioned Logical Identifier (LIDVID) – The concatenation of a logical identifier with a version identifier, providing a unique identifier for each version of product.

Metadata – Data about data. For example, a ‘description object’ contains information (metadata) about an ‘object.’

Object – A single instance of a class defined in the PDS Information Model.

PDS Information Model – The set of rules governing the structure and content of PDS metadata. While the Information Model (IM) has been implemented in XML for PDS4, the model itself is implementation independent.

Product – One or more labeled objects (digital, non-digital, or both) grouped together and having a single PDS-unique identifier. In the PDS4 implementation, if a product consists of multiple objects, their descriptions are combined into a single XML label. Although it may be possible to locate individual objects within PDS (and to find specific bit strings within digital objects), PDS4 defines ‘products’ to be the smallest granular unit of addressable data within its complete holdings.

RCE (Rover Compute Element) – One of the two redundant flight computers on Mars 2020. Either RCE (A or B) may be active at any given time.

Registry – A data base that provides services for sharing content and metadata.

XML schema – The definition of an XML document, specifying required and optional XML elements, their order, and parent-child relationships.

XML Schematron – A set of rules used to validate an XML document.

Version Identifier (VID) – Consist of major and minor components separated by a “.” (M.n), which identify a specific version of a product.

1 Overview

1.1 Purpose and Scope

This Software Interface Specification (SIS) describes the format and content of the Mars 2020 Scanning Habitable Environments with Raman & Luminescence for Organics and Chemicals (SHERLOC) Planetary Data System (PDS) data archive bundle in which data products, documentation, and supporting material are stored. This document is intended for the scientists who will analyze the data, including those associated with the project and those in the general planetary science community.

1.2 Contents

This SIS describes the organization, identification, and labeling of Mars 2020 SHERLOC raw and derived products. For details about these products, including how the instrument acquires data and how the data are processed, see the Mars 2020 SHERLOC EDR Data Product SIS [7] and RDR Data Product SIS [8].

1.3 Applicable Documents

- [1] PDS4 Concepts Document, version 1.16.0.0, April 21, 2021, <https://pds.nasa.gov/datastandards/documents/concepts/>.
- [2] Planetary Data System Standards Reference, version 1.16.0.0, April 21, 2021, <https://pds.nasa.gov/datastandards/documents/sr/>.
- [3] Planetary Data System Data Provider's Handbook, 1.16.0.0, April 21, 2021, <https://pds.nasa.gov/datastandards/documents/dph/>.
- [4] PDS4 Common Data Dictionary, Abridged, version 1.16.0.0, April 21, 2021, <https://pds.nasa.gov/datastandards/documents/dd/>.
- [5] PDS4 Information Model Specification, version 1.16.0.0, April 21, 2021, <https://pds.nasa.gov/datastandards/documents/im/>.
- [6] Mars 2020 Project Archive Generation, Validation and Transfer Plan, JPL D-95520, December 8, 2015.
- [7] Mars 2020 (M2020) Software Interface Specification: SHERLOC Spectroscopy Instrument Experiment Data Record (EDR) Data Products, JPL D-99965, December 15, 2020.
- [8] Mars 2020 Scanning Habitable Environments with Raman & Luminescence for Organics and Chemicals (SHERLOC) Reduced Data Record (RDR) Software Interface Specification, JPL D-104819, December 16, 2020.
- [9] Mars 2020 Software Interface Specification Camera Instrument Data Products, JPL D-99960, December 21, 2020.
- [10] Farley, K.A., K.H. Williford, K.M. Stack, et al., Mars 2020 Mission Overview, Space Sci. Rev 216, 142, 2020, doi:10.1007/s11214-020-00762-y.
- [11] Bhartia, R., L.W. Beegle, L. DeFlores, et al., Perseverance's Scanning Habitable Environments with Raman and Luminescence for Organics and Chemicals (SHERLOC) Investigation, Space Sci. Rev., 2021.

PDS4 is the name of the current PDS archive standard, described in Documents [1] through [5]. The PDS4 Information Model and Documents [1] through [5] are subject to periodic revision. The most recent

versions may be found at <https://pds.nasa.gov/datastandards/documents/>. The PDS4 products specified in this SIS have been designed based on the versions current at the time, which are those listed above. Data products will be archived using the version of the PDS Information Model that is current at the time the products are submitted to peer review. Peer-reviewed products do not need to be revised to incorporate subsequent changes in the Information Model.

1.4 Audience

This SIS is intended to be used both by the instrument team in generating the archive and by data users wishing to understand the format and content of the archive. Typically, these individuals would include scientists, data analysts, and software engineers.

1.5 Mars 2020 Mission

The Mars 2020 spacecraft launched in July of 2020 and placed the Perseverance Rover on the surface of Mars in Jezero Crater on February 18, 2021. The goal of the mission is to seek signs of life and to collect rock and soil samples for a future return to Earth. The rover will explore the landing site and acquire imaging, spectroscopy, and other measurements to characterize Martian soils, rocks, atmosphere, and other aspects of the environment. The rover carries seven scientific instruments and a sample acquisition and caching system. The various payload elements will be used as an integrated suite of tools to characterize the local geology, to study particular rock and soil targets, to characterize the local environment, and to acquire and cache selected rock and soil samples. The prime mission for the rover is expected to be 836 sols (approximately 2.5 Earth years), with the possibility of an extended mission of unknown duration after that. For additional information see the Mars 2020 mission paper [9].

1.6 SHERLOC Description

SHERLOC is an arm-mounted, deep ultraviolet (DUV) native fluorescence and resonance Raman spectrometer and imager. SHERLOC uses a DUV laser to illuminate the target and generate the characteristic Raman and fluorescence spectral response.

SHERLOC contains two imaging subsystems: Autofocus and Context Imager (ACI) and Wide Angle Topographic Sensor for Operations and eNginEering (WATSON). The DUV laser is co-boresighted to the ACI and an integrated scanning optical system that allows correlation of spectral signatures to surface textures, morphology and visible features. Through the use of an internal scanning mirror, autofocusing lens, and a depth of focus, the SHERLOC laser spot can be systematically scanned over the spectroscopy map area with a fine-scale spatial sampling on natural or abraded surfaces and borehole interior walls, without further arm movement. Through the use of the context imager, SHERLOC's data products will be combined with the other instruments on the Mars 2020 payload. This allows greatly increased scientific analysis by bearing down the entire scientific measurements on a single sample to assess the habitability of ancient environments and search for potential biosignatures preserved within the geologic record to select high-priority samples for caching.

Details of the SHERLOC instrument may be found in the SHERLOC EDR Data Product SIS [7].

2 SHERLOC Data Products

2.1 Data Product Overview

Mars 2020 SHERLOC data products in this bundle consist of raw and derived data. Table 1 shows a summary of all SHERLOC data product types. Data file formats are covered in Section 4 of this document. See the SHERLOC EDR Data Product SIS [7] and the SHERLOC RDR Data Product SIS [8] for details. The SHERLOC Process Flags column refers to characters 45 to 51 in the data product file name; see Table 4.

Table 1 All SHERLOC Data Product Types

Product Name	SHERLOC Product Type	SHERLOC Process Flags	Processing Level	PDS4 Bundle and Collection, with prefix urn:nasa:pds:	Description	PDS Data Type and File Name Extension
Spectroscopy EDR	ESH	n/a	raw	mars2020_sherloc: data_raw	Spectroscopy state of health in science mode	Table_Delimited, .CSV
Spectroscopy EDR	ECH	n/a	raw	mars2020_sherloc: data_raw	Spectroscopy state of health in calibration mode	Table_Delimited, .CSV
Spectroscopy EDR	EXH	n/a	raw	mars2020_sherloc: data_raw	Scanner position array metadata	Table_Delimited, .CSV
Spectroscopy EDR	ERA	n/a	raw	mars2020_sherloc: data_raw	Full resolution active spectra in science mode	Table_Delimited, .CSV
Spectroscopy EDR	ERB	n/a	raw	mars2020_sherloc: data_raw	Full resolution background (dark) spectra in science mode	Table_Delimited, .CSV
Spectroscopy EDR	ECA	n/a	raw	mars2020_sherloc: data_raw	Full resolution active spectra in calibration mode	Table_Delimited, .CSV
Spectroscopy EDR	ECB	n/a	raw	mars2020_sherloc: data_raw	Full resolution background (dark) spectra in calibration mode	Table_Delimited, .CSV
Spectroscopy EDR	EPA	n/a	raw	mars2020_sherloc: data_raw	Photodiode array in science mode	Table_Delimited, .CSV
Spectroscopy EDR	EPA	*****p*	raw	mars2020_sherloc: data_raw	Photodiode array in process_data mode	Table_Delimited, .CSV
Spectroscopy EDR	ESP	n/a	raw	mars2020_sherloc: data_raw	Scanner position array in science mode	Table_Delimited, .CSV
Spectroscopy EDR	ESP	*****p*	raw	mars2020_sherloc: data_raw	Scanner position array in process_data mode	Table_Delimited, .CSV
Spectroscopy EDR	ERP	*****p*	raw	mars2020_sherloc: data_raw	Spectroscopy and Photodiode data in process_data mode	Table_Delimited, .CSV
Spectroscopy RDR	RCA	****C**	calibrated	mars2020_sherloc: data_intermediate	Cosmic ray corrected active spectra in calibration mode	Table_Delimited, .CSV

Product Name	SHERLOC Product Type	SHERLOC Process Flags	Processing Level	PDS4 Bundle and Collection, with prefix urn:nasa:pds:	Description	PDS Data Type and File Name Extension
Spectroscopy RDR	RCA	****Z**	calibrated	mars2020_sherloc: data_intermediate	No Cosmic ray corrected active spectra in calibration mode	Table_Delimited, .CSV
Spectroscopy RDR	RAC	****C**	calibrated	mars2020_sherloc: data_intermediate	Cosmic ray active histogram in calibration mode	Table_Delimited, .CSV
Spectroscopy RDR	RCB	****C**	calibrated	mars2020_sherloc: data_intermediate	Cosmic ray corrected dark spectra in calibration mode	Table_Delimited, .CSV
Spectroscopy RDR	RCB	****Z**	calibrated	mars2020_sherloc: data_intermediate	No Cosmic ray corrected dark spectra in calibration mode	Table_Delimited, .CSV
Spectroscopy RDR	RBC	****C**	calibrated	mars2020_sherloc: data_intermediate	Cosmic ray dark histogram in calibration mode	Table_Delimited, .CSV
Spectroscopy RDR	RRA	****C**	calibrated	mars2020_sherloc: data_intermediate	Cosmic ray corrected active spectra in science mode	Table_Delimited, .CSV
Spectroscopy RDR	RRA	****Z**	calibrated	mars2020_sherloc: data_intermediate	No Cosmic ray corrected active spectra in science mode	Table_Delimited, .CSV
Spectroscopy RDR	RAR	****C**	calibrated	mars2020_sherloc: data_intermediate	Cosmic ray active histogram in science mode	Table_Delimited, .CSV
Spectroscopy RDR	RRB	****C**	calibrated	mars2020_sherloc: data_intermediate	Cosmic ray corrected dark spectra in science mode	Table_Delimited, .CSV
Spectroscopy RDR	RRB	****Z**	calibrated	mars2020_sherloc: data_intermediate	No Cosmic ray corrected dark spectra in science mode	Table_Delimited, .CSV
Spectroscopy RDR	RBR	****C**	calibrated	mars2020_sherloc: data_intermediate	Cosmic ray dark spectra in science mode	Table_Delimited, .CSV
Spectroscopy RDR	RRS	W*****	calibrated	mars2020_sherloc: data_intermediate	Dark-subtracted spectra in science mode. Data is processed by ground-derived algorithm, Wavelength Calibration.	Table_Delimited, .CSV
Spectroscopy RDR	RRS	*****G*	calibrated	mars2020_sherloc: data_intermediate	Dark-subtracted spectra in science mode. Data is processed by ground-derived algorithm, Gain Correction	Table_Delimited, .CSV
Spectroscopy RDR	RRS	*****N	derived	mars2020_sherloc: data_processed	Dark-subtracted spectra in science mode. Data is processed by ground-derived algorithm, Laser Normalization	Table_Delimited, .CSV
Spectroscopy RDR	RRS	*****Z	derived	mars2020_sherloc: data_processed	Dark-subtracted spectra in science mode. Data is not processed by ground-derived algorithm, Laser Normalization	Table_Delimited, .CSV

Product Name	SHERLOC Product Type	SHERLOC Process Flags	Processing Level	PDS4 Bundle and Collection, with prefix urn:nasa:pds:	Description	PDS Data Type and File Name Extension
Spectroscopy RDR	RCS	W*****	calibrated	mars2020_sherloc: data_processed	Dark-subtracted spectra in calibration mode. Data is processed by ground-derived algorithm, Wavelength Calibration.	Table_Delimited, .CSV
Spectroscopy RDR	RCS	*****G*	calibrated	mars2020_sherloc: data_processed	Dark-subtracted spectra in calibration mode. Data is processed by ground-derived algorithm, Gain Correction	Table_Delimited, .CSV
Spectroscopy RDR	RCS	*****N	derived	mars2020_sherloc: data_processed	Dark-subtracted spectra in calibration mode. Data is processed by ground-derived algorithm, Laser Normalization	Table_Delimited, .CSV
Spectroscopy RDR	RCS	*****Z	derived	mars2020_sherloc: data_processed	Dark-subtracted spectra in calibration mode. Data is not processed by ground-derived algorithm, Laser Normalization	Table_Delimited, .CSV
Spectroscopy RDR	RCC	n/a	derived	mars2020_sherloc: data_processed	Peak fit results to prominent peaks from calibration spectral data, trended over time in calibration mode	Table_Delimited, .CSV
Spectroscopy RDR	RRA	*****P*	calibrated	mars2020_sherloc: data_processed	No dark-subtracted spectra in process_data mode	Table_Delimited, .CSV
Spectroscopy RDR	RRS	*****P*	calibrated	mars2020_sherloc: data_processed	Dark-subtracted spectra in process_data mode	Table_Delimited, .CSV
Spectroscopy RDR	RRA	B*****P*	calibrated	mars2020_sherloc: data_processed	Dark frame spectra in process_data mode	Table_Delimited, .CSV
Spectroscopy RDR	RRA	****CP*	calibrated	mars2020_sherloc: data_processed	No dark-subtracted spectra with cosmic rays removed (on-board algorithm) in process_data mode	Table_Delimited, .CSV
Spectroscopy RDR	RRS	****CP*	calibrated	mars2020_sherloc: data_processed	Dark-subtracted spectra with cosmic rays removed (on-board algorithm) in process_data mode	Table_Delimited, .CSV
Spectroscopy RDR	RRA	****ZP*	calibrated	mars2020_sherloc: data_processed	No dark-subtracted spectra with no cosmic rays removed (on-board algorithm) in process_data mode	Table_Delimited, .CSV
Spectroscopy RDR	RRS	****ZP*	calibrated	mars2020_sherloc: data_processed	Dark-subtracted spectra with no cosmic rays removed (on-board algorithm) in process_data mode	Table_Delimited, .CSV
Spectroscopy RDR	RRA	*****PN	derived	mars2020_sherloc: data_processed	Non dark-subtracted spectra with laser normalization applied (on-board algorithm) in process_data mode	Table_Delimited, .CSV
Spectroscopy RDR	RRS	*****PN	derived	mars2020_sherloc: data_processed	Dark-subtracted spectra with laser normalization applied (on-board algorithm) in process_data mode	Table_Delimited, .CSV
Spectroscopy RDR	RRA	*****PZ	calibrated	mars2020_sherloc: data_processed	Non dark-subtracted spectra with no laser normalization applied (on-board algorithm) in process_data mode	Table_Delimited, .CSV
Spectroscopy RDR	RRS	*****PZ	calibrated	mars2020_sherloc: data_processed	Dark-subtracted spectra with no laser normalization applied (on-board algorithm) in process_data mode	Table_Delimited, .CSV

Product Name	SHERLOC Product Type	SHERLOC Process Flags	Processing Level	PDS4 Bundle and Collection, with prefix urn:nasa:pds:	Description	PDS Data Type and File Name Extension
Spectroscopy RDR	RMx	n/a	derived	mars2020_sherloc: data_processed	The integrated intensity for each laser shot using the spectral region in science mode. Each region could be tailored and defined by the ground processing, up to 34 regions. (x = 0-9, A-Z but not O)	Table_Delimited, .CSV
Spectroscopy RDR	RMx	*****p*	derived	mars2020_sherloc: data_processed	The integrated intensity for each laser shot using the spectral region in process_data mode. Each region could be tailored and defined by the ground processing, up to 34 regions. (x = 0-9, A-Z but not O)	Table_Delimited, .CSV
Spectroscopy RDR	RLI	n/a	calibrated	mars2020_sherloc: data_processed	The average laser photodiode intensity at each position in science mode	Table_Delimited, .CSV
Spectroscopy RDR	RLI	*****p*	calibrated	mars2020_sherloc: data_processed	The average laser photodiode intensity at each position in process_data mode	Table_Delimited, .CSV
Spectroscopy RDR	RLS	n/a	calibrated	mars2020_sherloc: data_processed	The line and sample number for each laser shot position on each ACI image, corrected to account for robotic arm drift in science mode	Table_Delimited, .CSV
Spectroscopy RDR	RLS	*****p*	calibrated	mars2020_sherloc: data_processed	The line and sample number for each laser shot position on each ACI image, corrected to account for robotic arm drift in process_data mode	Table_Delimited, .CSV
Spectroscopy RDR	RMO	n/a	derived	mars2020_sherloc: data_processed	A summary of image name, laser shot positions, and integrated intensity for each wavelength range at each point.	Table_Delimited, .CSV
Spectroscopy RDR	RMO	*****p*	derived	mars2020_sherloc: data_processed	A summary of image name, laser shot positions, and integrated intensity for each wavelength range at each point.	Table_Delimited, .CSV
ACI EDR			raw	mars2020_sherloc: data_aci	Raw image from ACI camera	Array_2D_Image, .IMG
ACI RDR			calibrated	mars2020_sherloc: data_aci	Calibrated image from ACI camera	Array_2D_Image, .IMG
WATSON EDR			raw	mars2020_sherloc: data_watson	Raw image from WATSON camera	Array_2D_Image, .IMG
WATSON RDR			calibrated	mars2020_sherloc: data_watson	Calibrated image from WATSON camera	Array_2D_Image, .IMG
EDR SIS	n/a		document	mars2020_sherloc: document	Description of EDR contents and format	Document, .PDF
RDR SIS	n/a		document	mars2020_sherloc: document	Description of RDR contents and format	Document, .PDF

Product Name	SHERLOC Product Type	SHERLOC Process Flags	Processing Level	PDS4 Bundle and Collection, with prefix urn:nasa:pds:	Description	PDS Data Type and File Name Extension
Bundle SIS	n/a		document	mars2020_sherloc:document	Description of mars2020_pixl bundle organization	Document, .PDF

2.2 Data Processing Levels

Data processing levels mentioned in this SIS refer to the PDS4 processing levels described in Table 2. The lowest processing level archived in PDS is “raw” as described in the table.

Table 2 Data Processing Level Definitions

PDS4 processing level	PDS4 processing level description
Raw	Original data from an experiment. If compression, reformatting, packetization, or other translation has been applied to facilitate data transmission or storage, those processes are reversed so that the archived data are in a PDS approved archive format. Often called EDRs (Experimental Data Records).
Partially Processed	Data that have been processed beyond the raw stage but which have not yet reached calibrated status. These and more highly processed products are often called RDRs (Reduced Data Records).
Calibrated	Data converted to physical units, which makes values independent of the experiment.
Derived	Results that have been distilled from one or more calibrated data products (for example, maps, gravity or magnetic fields, or ring particle size distributions). Supplementary data, such as calibration tables or tables of viewing geometry, used to interpret observational data should also be classified as ‘derived’ data if not easily matched to one of the other three categories.

3 SHERLOC Archive Organization

This section describes the organization of the SHERLOC archive according to the PDS4 Information Model [5].

3.1 The SHERLOC Bundle

The highest level of organization for a PDS archive is the bundle. A bundle is a set of one or more related collections that may be of different types. A collection is a set of one or more related basic products that are typically all of the same type. Bundles and collections are logical structures, not necessarily tied to any physical directory structure or organization. Figure 1 illustrates the relationships among bundles, collections, and products. A product consists of one or more objects (e.g., a table of data, an image, or a document) described by a label.

The complete SHERLOC archive is organized into one bundle. The bundle’s PDS Logical Identifier (LID) is **urn:nasa:pds:mars2020_sherloc**.

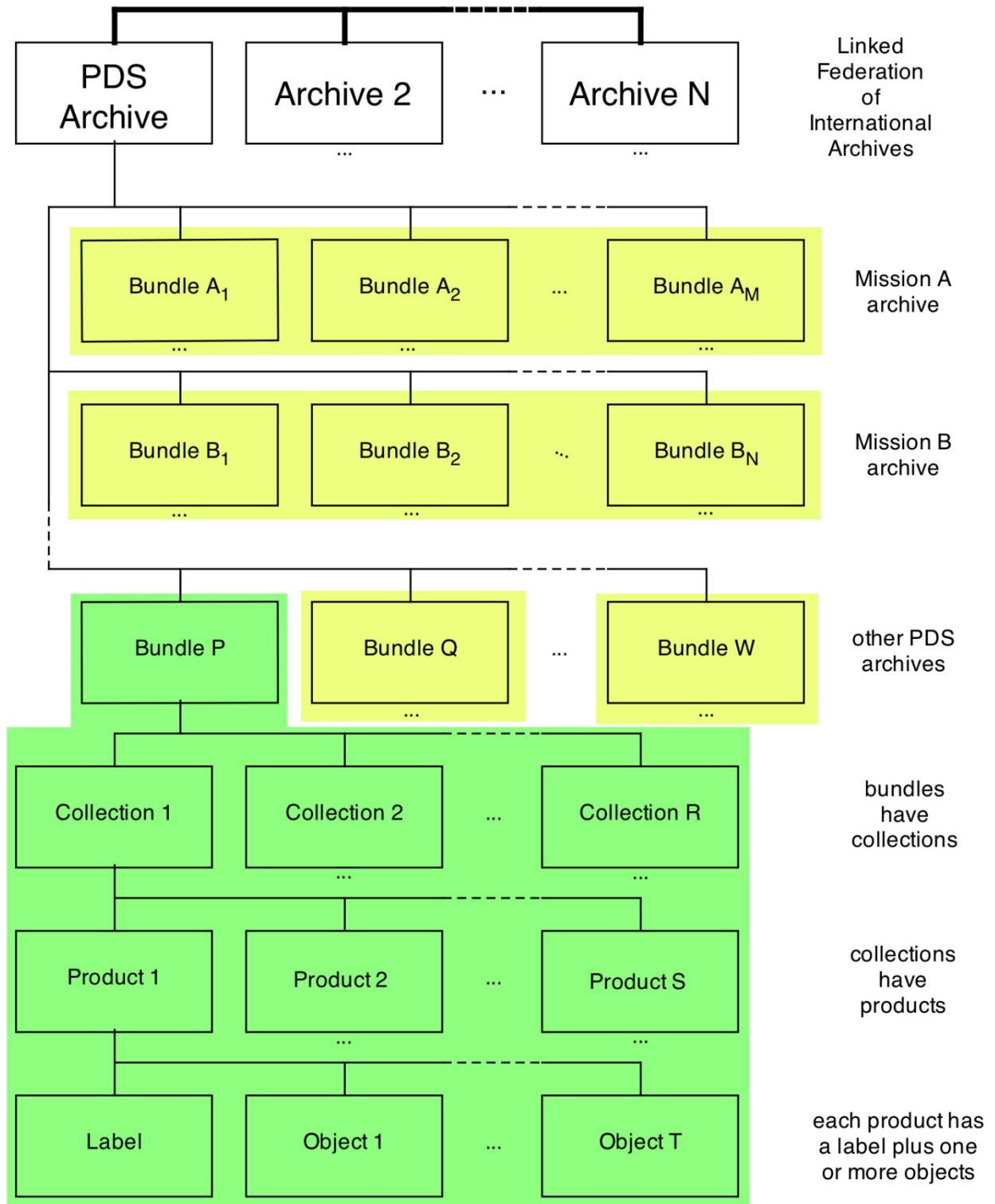


Figure 1 PDS bundles, collections, and products.

3.2 SHERLOC Collections

The SHERLOC bundle includes the following collections of data products and documents.

Table 3 Collections in the SHERLOC Bundle

Collection Logical Identifier	Collection Type	Contents
<code>urn:nasa:pds:mars2020_sherloc:data_raw</code>	Data	Raw (EDR) non-image science data products
<code>urn:nasa:pds:mars2020_sherloc:data_intermediate</code>	Data	Partially processed (RDR) non-image science data products
<code>urn:nasa:pds:mars2020_sherloc:data_processed</code>	Data	Calibrated and derived (RDR) non-image science data products
<code>urn:nasa:pds:mars2020_sherloc:data_aci</code>	Data	Raw and calibrated image data products from the ACI camera
<code>urn:nasa:pds:mars2020_sherloc:data_watson</code>	Data	Raw and calibrated image data products from the WATSON camera
<code>urn:nasa:pds:mars2020_sherloc:document</code>	Document	Documentation, including this SIS and the EDR and RDR Data Product SISes

3.3 SHERLOC Data Organization

The SHERLOC bundle is organized as a single directory tree with a subdirectory for each collection. Data subdirectories are further subdivided by sol (Mars day). See Appendix B for a diagram of the full directory structure of the bundle.

3.4 SHERLOC Product Identification and Naming

A SHERLOC data product consists of one or more digital objects in one file, accompanied by a PDS label file. A table of data is an example of a digital object. The PDS label provides identification and other metadata for the data file. The PDS label typically has the same name as the file it describes, except that it has the extension ".xml". See section 3.5 for more information about PDS labels.

In addition to data products the bundle includes a collection of document products, which also have PDS labels. The document collection includes the SHERLOC EDR Data Product Software Interface Specification (SIS), the SHERLOC RDR Data Product SIS, and the SHERLOC Bundle SIS (this document).

Finally, the collections and the bundle themselves are considered products in PDS, and therefore have their own labels. The bundle label is in the root directory of the archive and is named **bundle_sherloc.xml**. The label for a collection is found in the directory for the collection, and is named **collection_<something>.xml**. For example, the label for the **data_raw** collection is the file **collection_data_raw.xml** in the directory **data_raw**. Each collection has an inventory table that lists the products in the collection, e.g. **collection_data_raw_inventory.csv**. For example, the first few lines of **collection_data_raw_inventory.csv** might look like this:

```
P,urn:nasa:pds:mars2020_sherloc:data_raw:ss__d004t0631435218_975epa__0010256src01002_0__p_j:::1.0
P,urn:nasa:pds:mars2020_sherloc:data_raw:ss__d004t0631436301_370ecb__0010256src01002_0__j:::1.0
P,urn:nasa:pds:mars2020_sherloc:data_raw:ss__d004t0631446724_619esh__0010256src01001_0__j:::1.0
```

The first field is either “P” or “S”, indicating whether the product is a Primary or Secondary member of the collection. The second field is the complete logical identifier of the product.

3.4.1 Logical Identifiers

Every product in PDS is assigned a Logical Identifier (LID) that allows it to be uniquely identified across the system. Each product also has a Version Identifier (VID) that allows different versions of a specific product to be referenced uniquely. A product’s LID and VID are defined as separate attributes in the product label. For convenience they may be combined in a single string called a LIDVID, with two colons between the LID and the VID. If a particular version of a product is desired, the LIDVID should be used; otherwise the LID alone should be used with the understanding that it refers to the latest version of the product. Example LIDs are given in the following section.

LIDs and VIDs are assigned by PDS and are formed according to the conventions described in the following sections. More information on LIDs and VIDs may be found in Section 6D of the PDS Standards Reference [2] and in Chapter 5 of the Data Providers' Handbook [3].

3.4.1.1 LID Formation

LIDs take the form of a Uniform Resource Name (URN). LIDs are restricted to ASCII lower case letters, digits, dash, underscore, and period. Colons are used to separate prescribed components of the LID. Within one of these prescribed components, the dash, underscore, or period may be used as separators. LIDs are limited in length to 255 characters.

Mars 2020 LIDs are formed according to the following conventions:

- Bundle LIDs are formed by appending a bundle-specific ID to the PDS base ID:

```
urn:nasa:pds:<bundle ID>
```

Example: **urn:nasa:pds:mars2020_sherloc**

The bundle ID must be unique across all bundles archived with the PDS.

- Collection LIDs are formed by appending a collection-specific ID to the collection’s parent bundle LID:

```
urn:nasa:pds:<bundle ID>:<collection ID>
```

Example: **urn:nasa:pds:mars2020_sherloc:data_processed**

The collection ID must be unique across the bundle. Collection IDs correspond to the collection type (e.g. “browse”, “data”, “document”, etc.). Additional descriptive information may be appended to the collection type (e.g. “data_raw”, “data_calibrated”, etc.).

- Basic product LIDs are formed by appending a product-specific ID to the product’s collection LID:

```
urn:nasa:pds:<bundle ID>:<collection ID>:<product ID>
```

Example:

```
urn:nasa:pds:mars2020_sherloc:data_processed:ss__0004_0667298185_056rcc__0010052src10002w0__cgzj01
```

- The product LID must be unique across the collection. For SHERLOC RDR data products, the product LID is the same as the lowercased data file name without the extension. For SHERLOC EDR data products, the product LID is the same as the lowercased data file name with the extension.

3.4.1.2 VID Formation

Product Version IDs consist of major and minor components separated by a “.” (M.n). Both components of the VID are integer values.

For SHERLOC EDRs, the major component is set based on the operations pipeline version ID assigned to the product. This value is extracted from the last two characters of the ops product filename. For SHERLOC EDR products the minor version is always “0”. Thus the complete VID for all SHERLOC EDR products is “M.0”. As not all internal product versions are released to PDS, “M” does not necessarily begin at 1, and revisions in the PDS archive may skip version numbers. This versioning scheme does not follow the PDS standard for versioning, which requires versions to begin at 1.0 and increment sequentially.

For SHERLOC RDRs, versioning follows the rules in the PDS Standards Reference [2]. The major component is initialized to a value of “1”, and the minor component is initialized to a value of “0”. The minor component resets to “0” when the major component is incremented.

Example of a complete LIDVID for a SHERLOC RDR data product, version 1.0:

urn:nasa:pds:mars2020_sherloc:data_processed:ss_0004_0667298185_056rcc_0010052src10002w0_cgzi01::1.0

3.4.2 File Naming Convention

Figure 2 and Table 4 illustrate the file naming convention for SHERLOC non-imaging data products. This convention is used for most Mars 2020 data files, and some fields do not apply to SHERLOC. More details about this naming convention are found in the SHERLOC EDR Data Product SIS [7].

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58
INSTRUMENT	CONFIG	SPECIAL FLAG	SOL NUMBER	VENUE	SCLK															-	SUB-SCLK	PROD TYPE	GEOMETRY THUMBNAIL	SITE	DRIVE	SEQUENCE/RTT										INST. SPECIFIC	PRODUCER	VERSION	EXT																		

Figure 2 Mars 2020 File Naming Scheme

Table 4 SHERLOC File Name Components

Characters	Contents	Description						
1-2	Instrument	SS: SHERLOC Spectroscopy SC: SHERLOC ACI SI: SHERLOC WATSON						
3	Configuration	“_” for SHERLOC spectroscopy and engineering For image products, cover and LED state: <table border="1" style="margin-left: 20px;"> <tr> <td>0</td> <td>Cover closed, LEDs off</td> </tr> <tr> <td>1</td> <td>Cover open, LEDs off</td> </tr> <tr> <td>2</td> <td>Cover closed, LEDs on</td> </tr> </table>	0	Cover closed, LEDs off	1	Cover open, LEDs off	2	Cover closed, LEDs on
0	Cover closed, LEDs off							
1	Cover open, LEDs off							
2	Cover closed, LEDs on							

		4	Cover closed, LEDs on												
4	Special Processing Flag	“_” indicates none “2” indicates a 2D read-out of the detector was performed (ERA, ERB, ECA, ECB only)													
5-8	Sol	Sol number for flight surface data; see EDR SIS for other uses													
9	Mission venue	“_” for Flight Mode													
10-19	Spacecraft Clock Count (SCLK)	10-integer spacecraft clock count in seconds													
20	Underscore	Always “_” for readability													
21-23	Fractional SCLK	3-digit spacecraft clock count fractional seconds													
24-26	Product Type	See SHERLOC Product Type in Table 1													
27	Geometry	“_” for non-image products. For image products denote for Geometry for WATSON and ACI: “_” indicates raw (non-linearized) geometry “L” indicates product has been linearized with nominal stereo partner “A” indicates product has been linearized with actual stereo partner													
28	Thumbnail	“_” for non-image products “T”: Product is a thumbnail images “N”: Product is not a thumbnail image													
29-31	Site	Site identifier from Rover Motion Counter													
32-35	Drive	Drive identifier (position within a site location)													
36-44	Sequence ID	Identifier indicating the command sequence from which the data were acquired													
Start of Instrument Specific Identifier Fork															
A. Image Products only (Char 45-51)															
45-48	Camera-specific identifier	For image products: Snnn where nnn is an observation number													
49	Downsample resolution	“_” for non-image data products For image products: $Resolution = 2^n \times 2^n$ <table border="1" data-bbox="862 1396 1252 1696"> <thead> <tr> <th>Valid values</th> <th>Resolution</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>1x1</td> </tr> <tr> <td>1</td> <td>2x2</td> </tr> <tr> <td>2</td> <td>4x4</td> </tr> <tr> <td>3</td> <td>8x8</td> </tr> <tr> <td>...</td> <td>...</td> </tr> </tbody> </table>		Valid values	Resolution	0	1x1	1	2x2	2	4x4	3	8x8
Valid values	Resolution														
0	1x1														
1	2x2														
2	4x4														
3	8x8														
...	...														
50-51	Compression	“_” for non-image data products For image products: <table border="1" data-bbox="883 1759 1417 1892"> <thead> <tr> <th>Type</th> <th>Valid values</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>JPEG (lossy)</td> <td>00 01-99</td> <td>Thumbnail Jpeg quality level</td> </tr> </tbody> </table>		Type	Valid values	Description	JPEG (lossy)	00 01-99	Thumbnail Jpeg quality level						
Type	Valid values	Description													
JPEG (lossy)	00 01-99	Thumbnail Jpeg quality level													

		A0	Jpeg quality level 100																				
	ICER (lossy)	I1, I2, ..., I8 I9	1 bpp, 2 bpp, ..., 8 bpp Anything higher than 8 bpp																				
	Lossless	LI LL LM LU	ICER LOCO Malin Uncompressed																				
B. Spectroscopy Products "Process Flags" only (Char 45-51)																							
45	Instrument specific identifier	"W" : Wavelength Correction algorithm performed "B" : On-board process_data product is background-subtracted "_ " : No Wavelength Correction algorithm performed																					
46	Instrument specific identifier.	Spectroscopy experiment ID designation.																					
		<table border="1"> <thead> <tr> <th>Values</th> <th>Range</th> </tr> </thead> <tbody> <tr> <td>0, 1, ..., 9</td> <td>0 thru 9</td> </tr> <tr> <td>A, B, ..., Z</td> <td>10 thru 35</td> </tr> </tbody> </table>	Values	Range	0, 1, ..., 9	0 thru 9	A, B, ..., Z	10 thru 35															
Values	Range																						
0, 1, ..., 9	0 thru 9																						
A, B, ..., Z	10 thru 35																						
47-48	Instrument specific identifier	The valid value is the total number of ACI images acquired, in relation to the ExpID. The ACI image number can ranges between "00" thru "99".																					
49	Instrument specific identifier	"C" : Cosmic Ray Correction algorithm performed "Z" : Cosmic Ray Correction has been skipped "_ " : No Cosmic Ray Correction algorithm performed																					
50	Instrument specific identifier	"G" : Gain Correction algorithm performed "P" : On-board process_data algorithm has been applied "_ " : No Gain Correction algorithm performed																					
51	Instrument specific identifier	"N" : Normalization algorithm performed "Z" : Laser Normalization has been skipped "_ " : No Normalization Correction algorithm performed																					
End of Instrument Specific Identifier Fork																							
52	Provider	Provider institution ID "J": IDS at JPL "P": SHERLOC Principal Investigator at JPL																					
53-54	Product Version	Product version number. Increments by one whenever a previously generated file with an otherwise identical filename exists.																					
		<table border="1"> <thead> <tr> <th>Values</th> <th>Range</th> </tr> </thead> <tbody> <tr> <td>00, 01, 02 ..., 99</td> <td>0 thru 99</td> </tr> <tr> <td>A0, A1, ..., A9</td> <td>100 thru 109</td> </tr> <tr> <td>AA, AB, ..., AZ</td> <td>110 thru 135</td> </tr> <tr> <td>B0, B1, B2 ..., B9</td> <td>136 thru 145</td> </tr> <tr> <td>BA, BB, ..., BZ</td> <td>146 thru 171</td> </tr> <tr> <td>...</td> <td>...</td> </tr> <tr> <td>Z0, Z1, ..., Z9</td> <td>1000 thru 1009</td> </tr> <tr> <td>ZA, ZB, ..., ZZ</td> <td>1010 thru 1035</td> </tr> <tr> <td>_ _</td> <td>Value is out of range</td> </tr> </tbody> </table>	Values	Range	00, 01, 02 ..., 99	0 thru 99	A0, A1, ..., A9	100 thru 109	AA, AB, ..., AZ	110 thru 135	B0, B1, B2 ..., B9	136 thru 145	BA, BB, ..., BZ	146 thru 171	Z0, Z1, ..., Z9	1000 thru 1009	ZA, ZB, ..., ZZ	1010 thru 1035	_ _	Value is out of range	
Values	Range																						
00, 01, 02 ..., 99	0 thru 99																						
A0, A1, ..., A9	100 thru 109																						
AA, AB, ..., AZ	110 thru 135																						
B0, B1, B2 ..., B9	136 thru 145																						
BA, BB, ..., BZ	146 thru 171																						
...	...																						
Z0, Z1, ..., Z9	1000 thru 1009																						
ZA, ZB, ..., ZZ	1010 thru 1035																						
_ _	Value is out of range																						
55	Separator	Separator for filename and extension, always "."																					

56-58	File name extension	"DAT" : Binary table "CSV" : ASCII comma-separated-value text file "IMG" : Image data
-------	---------------------	---

The following are examples of SHERLOC data product file names.

Science RCA with Cosmic Ray Correction	ss__0001t0631434189_340rca__0010256srlc01001_008c__j01.csv
Science RCS with Cosmic Ray and Gain Corrections	ss__0001t0631434189_340rcs__0010256srlc01001_008cg_j01.csv
Science RCS with Cosmic Ray, Gain Correction, and Wavelength Corrections	ss__0001t0631434189_340rcs__0010256srlc01001w008cg_j01.csv
Science RRS with Cosmic Ray, Gain, Wavelength, and Laser Normalization Corrections.	ss__0001t0631434170_215rrs__0010256srlc01001w008cgnj01.csv
Process_data RRA with no Cosmic Ray nor Laser Normalization performed on-board.	ss__0002t0631435998_000rra__0010256srlc01002_003zpzj01.csv
Process_data RLI	ss__0002t0631435998_000rli__0010256srlc01002_003_p_j01.csv

SHERLOC data file names in the PDS archive are in lowercase characters.

3.5 PDS4 Labels

A typical PDS4 data product consists of a data object (e.g., a table) in one file and an accompanying label in a separate file. A product may have more than one data object in the data file, and it may have more than one data file, but it always has exactly one label in a file by itself.

PDS4 labels are ASCII text files written in the eXtensible Markup Language (XML). Typically a label has the same file name as the data file it describes but with the extension “.xml”. If the label describes more than one data file, it will have a slightly different file name, but always the extension “.xml”. If the data file also happens to be an XML file, its file name extension will be changed to avoid a conflict with the label file name.

Documents are also considered products and have accompanying PDS4 labels.

3.6 PDS4 Data Dictionaries

The structure and content of PDS4 labels conform to the PDS Information Model as embodied by the PDS Common Dictionary [4] and, as needed, additional mission-specific or discipline-specific data dictionaries. PDS dictionaries are written as XML schema and Schematron files, which are maintained at <https://pds.nasa.gov/datastandards/dictionaries/>. The PDS Data Provider’s Handbook explains the use of these schema and Schematron files [3]. In brief, the schema is the XML model that PDS4 labels must follow, and the Schematron is a set of validation rules that are applied to PDS4 labels.

At the beginning of every PDS4 label are statements listing the name and version of the PDS Common Dictionary and any other data dictionaries that are used in the label. The PDS Validate Tool

(<https://pds.nasa.gov/tools/about/>) is used by data providers and by PDS to ensure that the label conforms to the dictionary specifications, and that the label correctly describes the contents of the data file(s).

Table 5 lists the data dictionaries used in SHERLOC labels.

Table 5 PDS4 Dictionaries Used In SHERLOC Labels

Dictionary	File Name	Steward
PDS Common Dictionary	PDS4_PDS_1G00.*	PDS Engineering Node
Mars 2020 Mission Dictionary	PDS4_MARS2020_1G00_1000.*	PDS Geosciences Node
Geometry Discipline Dictionary	PDS4_GEOM_1G00_1930.*	PDS Geosciences Node
Processing Information Discipline Dictionary	PDS4_PROC_1G00_1210.*	PDS Cartography and Imaging Sciences Node
Mission Surface Discipline Dictionary	PDS4_MSN_SURFACE_1G00_1220.*	PDS Cartography and Imaging Sciences Node

***Versions current at time of first data release**

4 SHERLOC Product Formats

This section describes the formats of data and document product types in the SHERLOC bundle.

4.1 Data Product Formats

SHERLOC non-image data products are formatted as ASCII text comma-separated value (CSV) tables, according to the definitions of these structures in Section 4 of the PDS Standards Reference [2]. The formats are described at length in the SHERLOC Data Product SIS [7].

A SHERLOC non-image data file may contain one or more CSV tables. Some CSV tables are preceded by a header record.

SHERLOC image data products are formatted as PDS image (IMG) files. An IMG file is a binary array of values all of the same data type. Image products are described in the Mars 2020 Camera SIS [9].

4.2 Document Product Formats

Documents in this archive are in Portable Document File (PDF) format, specifically PDF/A (<https://www.loc.gov/preservation/digital/formats/fdd/fdd000125.shtml>), or in plain ASCII text in cases where no special formatting is required.

Appendix A Support Staff and Cognizant Persons

Name	Role	Institution	Email
Susan Slavney	PDS Node Representative	Washington University	susan.slavney@wustl.edu
Amy Chen	SHERLOC Representative	Jet Propulsion Laboratory	amy.c.chen@jpl.nasa.gov
Deborah Padgett	IDS Representative	Jet Propulsion Laboratory	deborah.l.padgett@jpl.nasa.gov

Appendix B Bundle Directory Structure

SHERLOC Bundle Root

```
|--- bundle_sherloc.xml
|--- readme.txt
|--- data_aci
    |--- collection_data_aci.xml
    |--- collection_data_aci_inventory.csv
|--- data_intermediate
    |--- collection_data_intermediate.xml
    |--- collection_data_intermediate_inventory.csv
    |--- sol_00001
        . . .
    |--- sol_nnnnn
|--- data_processed
    |--- collection_data_processed.xml
    |--- collection_data_processed_inventory.csv
    |--- sol_00001
        . . .
    |--- sol_nnnnn
|--- data_raw
    |--- collection_data_raw.xml
    |--- collection_data_raw_inventory.csv
    |--- sol_00001
        . . .
    |--- sol_nnnnn
|--- data_watson
    |--- collection_data_watson.xml
    |--- collection_data_watson_inventory.csv
|--- document
    |--- collection_document.xml
    |--- collection_document_inventory.csv
    |--- m20_sherloc_edr_dpo_structure.pdf, .xlsx, .xml
    |--- mars2020_sherloc_labels_sort_*.html, .pdf
    |--- sherloc_bundle_sis.pdf, .xml
    |--- sherloc_edr_sis.pdf, .xml
    |--- sherloc_rdr_sis.pdf, .xml
    |--- sherloc_release_notes.pdf, .xml
```

```
|--- sherloc_user_guide.pdf, .xml
```

The **data_aci** and **data_watson** collections contains SHERLOC ACI and WATSON images, which are primary members of the collection **urn:nasa:pds:mars2020_imgops:data_mcc_imgops**. They are listed by their LIDVIDs as secondary members in the files **collection_data_aci_inventory.csv** and **collection_data_watson_inventory.csv**. The LIDVIDs may be used to locate the images in their primary collection, using PDS search tools.