Scanning Habitable Environments with Raman & Luminescence for Organics and Chemicals (SHERLOC) Reduced Data Record (RDR) Software Interface Specification

Project Control Doc ID:
JPL D-104819

Project Repository (for life of mission):
TBA

Version: 1.12
Release Date: July 15, 2021

Authors:
Amy Chen, Kyle Uckert, Luca Cinquini

Custodian:
Kyle Uckert

Signatures

SHERLOC Generating Elements:

Science Data System (SDS)
Amy Chen

____
System Lead Date

Science Receiving Elements:

Deputy Project Scientist
Ken Williford

____
Deputy Project Scientist Date

Concurrence:

M2020 SHERLOC Investigation Scientist
Luther Beegle

____
Investigation Scientist Date

M2020 SHERLOC Instrument Scientist
Kevin Hand

____
Instrument Scientist Date

PDS Program Manager
Timothy McClanahan

____
Manager Date

PDS Geosciences Node Manager
Raymond Arvidson

____
Manager Date

Mars 2020 Project
Software Interface Specification

Scanning Habitable Environments with Raman & Luminescence for Organics and Chemicals (SHERLOC) Reduced Data Record (RDR)

JPL D-104819

Authors:
Amy Chen, Kyle Uckert, Luca Cinquini

Custodian:
Kyle Uckert

Paper copies of this document may not be current and should not be relied on for official purposes. The current version is on the project-internal repository at: TBA

CHANGE LOG

DATE	SECTIONS CHANGED	REVISION
2020-06-09	All	Draft
2020-08-04	All	v0.6
2020-10-28	All	v0.7
2020-12-16	All	v0.8
2020-12-17	Sections 3.1, 4.2.2, 4.4 and Appendix A– Corrected Table Figure references in the paragraphs. Sections 5.3.1 and 5.3.4 – Corrected Figure references in the paragraphs.	v0.81
2021-02-04	Revisions after initial PDS review	V0.9
2021-05-03	Revisions after PDS Peer Review feedbacks	V1.0
2021-06-25	reordered Sections 4.4.9 and 4.4.10 removed Track Changes comments reordered entries in Table 3 updated cosmic ray correction description (4.4.1, 4.4.2)	V1.1
2021-06-30	Updated validate link, updated (line, sample) description, removed ":" from Appendix B	V1.11
2021-07-15	Removed TBD in Table 7 Revised Appendix D	V1.12
2021-07-29	Removed Appendix C and D	

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS AND ABBREVIATIONS

ACI	Autofocus and Context Imager
ASCII	American Standard Code for Information Interchange
ATLO	Assembly, Test, Launch and Operations
CCD	Charge Coupled Device
CODMAC	Committee on Data Management and Computation
CSV	Comma-separated-value
DEA	Digital Electronics Assembly
DOY	Day of Year
DP	Data Product (telemetry)
EDR	Experiment Data Record
EHA	Engineering, Housekeeping & Accountability (EH&A)
EM	Engineering Model
EMD	Earth Metadata file (".emd")
ERT	Earth Received Time
FM	Flight Model
FOV	Field of View
FSW	Flight Software
FWHM	Full Width at Half Maximum
GDS	Ground Data System
ICD	Interface Control Document
IDS	Instrument Data Systems
JPL	Jet Propulsion Laboratory
LANL	Los Alamos National Laboratory
MIPL	Multimission Instrument Processing Laboratory
MS	Mission System
M2020	Mars 2020 Rover
MSSS	Malin Space Science Systems
NASA	National Aeronautics and Space Administration
OCS	Operations Cloud Store
ODL	Object Description Language
ORT	Operations Readiness test
PDS	Planetary Data System
PGE	Product Generation Executable
PRT	Platinum Resistance Thermometer
PSDD	Planetary Science Data Dictionary
RA	Robotic Arm

RCE	Rover Compute Element
RMC	Rover Motion Counter
RDR	Reduced Data Record
ROI	Region of Interest
SCCD	Spectrometer CCD
SCID	Spacecraft ID
SCLK	Spacecraft Clock
SCT	SHERLOC Calibration Target
SDS	Science Data System
SFDU	Specific Formatted Data Unit
SHERLOC	Scanning Habitable Environments with Raman and Luminescence for Organics and Chemicals
SIS	Software Interface Specification
SOAS	Science Operations Analysis Subsystem
SOH	State of Health
SOL	Mars Solar Day
SPICE	Spacecraft, Planet, Instrument, C-matrix, Events kernels
TBA	To Be Added
TBU	Testbed Unit
TDS	Telemetry Delivery Subsystem
WATSON	Wide Angle Topographic Sensor for Operations and eNgineering

1. Introduction

1.1 Purpose and Scope

The purpose of this Software Interface Specification (SIS) is to provide consumers of Mars 2020 (M2020) SHERLOC Reduced Data Record (RDR) data products with a detailed and informative description of the instrument data products.

The content in this document supports RDR data products generated by the SHERLOC instrument team. SHERLOC instrument comprises three subsystems:

Auto-focus Context Imager (ACI)
Deep ultra-violet (DUV) Raman and Luminescence Spectrometer
Wide Angle Topographic Sensor for Operations and eNgineering (WATSON)

In this document, the Experiment Data Record (EDR) data product refers to raw, uncalibrated and/or uncorrected data acquired by the three subsystems mentioned above. Imager EDRs, for both ACI and WATSON, may include image decompression if there was data product compression performed by the instrument onboard the Mars 2020 rover, *Perseverance*. The EDR descriptions of the ACI, WATSON and spectroscopy are documented separately by the M2020 mission system (MS) Instrument Data Subsystem (IDS) element, see Section 1.3 Ref [5] and Ref [6]. The IDS element may generate imager RDRs that are beyond the instrument team's scope or responsibility. The IDS element will be responsible to document those data products.

The Spectroscopy RDRs are derived from the uncalibrated and uncorrected spectroscopy EDRs. The SHERLOC instrument team performs post EDR downstream data processing to extract, interpret, and output various RDRs to support tactical and strategic rover surface operations needs. Not all RDRs generated by the instrument are intended to be archived. Appendix A provides the list of SHERLOC science data that are intended for archival.

The SHERLOC instrument team may generate derived data products from ACI and/or WATSON imaging data products to overlay with spectroscopy map products. Output data files of this type of derived imaging RDR will be written out as CSV files. These files will be archived.

1.2 Contents

This SIS describes how the SHERLOC instrument acquires its data and how the data are processed, formatted, labeled and uniquely identified. The document discusses standards used in generating the product and software that may be used to access the product. The data product structure and organization is described in sufficient detail to enable a user to read the product. Examples of product labels are provided. The different product types are described below in Section 4.4.

1.3 Constraint and Applicable Documents

This SIS is meant to be consistent with the contract negotiated between the M2020 Project and the SHERLOC Principal Investigators (PI/Co-PI). As this SIS governs the specification of SHERLOC data products used during M2020 surface operations, any proposed changes to this SIS may be impacted by all affected software subsystems observing this SIS in support of operations, e.g. SHERLOC SDS, MS/IDS, SOAS.

Product label keywords may be added to a future revision of this SIS. Therefore, it is recommended that software designed to process RDRs specified by this SIS should be robust to potential new keywords.

The SIS is governed by the following M2020 documents:

1. Mars Exploration Program Data Management, Rev. 4, June 15, 2011
2. Mars 2020 Project Archive Generation Validation and Transfer Plan, JPL D-95520
3. SHERLOC Functional Design Description, JPL D-95109
4. SHERLOC Science Team and PDS Geoscience Node Interface Control Document (ICD), Version 1.2, August 11, 2015
5. Mars 2020 Camera Instrument EDR/RDR Data Product Software Interface Specification (SIS), JPL D-99960
6. SHERLOC Spectroscopy EDR Data Products Software Interface Specification (SIS), JPL D-99965
7. M2020 Pointing, Positioning, Phasing and Coordinate Systems (PPPCS)
8. M2020 Surface Attitude, Positioning and Pointing (SAPP) Functional Design Description (FDD)

This SIS is also consistent with the following Planetary Data System documents:

9. Planetary Data System (PDS) PDS4 Information Model Specification, https://pds.nasa.gov/datastandards/documents/im/current/
10. Planetary Data System Standards Reference, https://pds.nasa.gov/datastandards/documents/sr/current/
11. PDS4 Data Dictionary, Abridged, https://pds.nasa.gov/datastandards/documents/dd/current/
12. PDS4 Common XML Schema and PDS4 Schematron, and other Schemas and Schematron files recognized in PDS4, https://pds.nasa.gov/datastandards/schema/released/

Other relevant documents include the following instrument description papers:

13. R. Bhartia, L. W. Beegle, K. Edgett, et al., The Mars 2020 Rover Scanning Habitable environments with Raman and Luminescence for Organics and Chemicals (SHERLOC) Investigation, Space Science Review, [tba]
14. K. Uckert, R. Bhartia, J. Michel, A Semi-Autonomous Method to Detect Cosmic Rays in Raman Hyperspectral Data Sets, Applied Spectroscopy, July 25, 2019 doi: 10.1177%2F0003702819850584

15. T. G. Graff, R, Bhartia, L. w. Beegle, et al., The Calibration Target for the Mars 2020 SHERLOC Instrument. *LPI*, (2132), p.2717.
16. L. W. Beegle, et al., SHERLOC: Scanning Habitable Environment with Raman and Luminescence for Organics and Chemicals. 2015 IEEE Aerospace Conference. IEE, 2015
17. M. Caffrey, K. Gasway, J. McGlown, J. Michel, A. Nelson, … & L. Beegle. The Processing Electronics and Detector of the Mars 2020 SHERLOC Instrument. 2020 IEEE Aerospace Conference (pp. 1-8).
18. K. Uckert, R. Bhartia, L. Beegle, et al. Calibration of the SHERLOC Deep UV Fluorescence/Raman Spectrometer on the Perseverance Rover. Applied Spectroscopy, *accepted*.
doi: 10.1177%2F00037028211013368

1.4 Relationships with Other Interfaces

Changes to SHERLOC EDR/RDR data product SIS document affect the following products, software and/or documents.

Table 1 Product and Software Interfaces to this SIS

Software Name	Type P = product S = software D = document	Owner
MIPL Database Schema	P	MIPL (JPL)
M2020 Camera Instrument EDRs	P	MIPL (JPL)
M2020 SHERLOC DUV Spectroscopy EDRs	P	MIPL (JPL)
Spectrum Unpack	S	SHERLOC (JPL)
Process Data Unpack	S	SHERLOC (JPL)
Gain Correction	S	SHERLOC (JPL)
Dark Subtraction	S	SHERLOC (JPL)
Cosmic Ray Correction	S	SHERLOC (JPL)
Wavelength Correction	S	SHERLOC (JPL)
Laser Photodiode Normalization	S	SHERLOC (JPL)
Calibration Comparison	S	SHERLOC (JPL)
XY Conversion	S	SHERLOC (JPL)
Map Conversion	S	SHERLOC (JPL)
Image Correction	S	SHERLOC (JPL)
Image Comparison	S	SHERLOC (JPL)
Image Map Overlays	S	SHERLOC (JPL)
Autolook Generator	S	SHERLOC (JPL)
Engineering Unpack	S	SHERLOC (JPL)
SHERLOC Autolook	P	SHERLOC (JPL)
SOH Report	P	SHERLOC (JPL)
SHERLOC Downlink Assessment JNB	S	SHERLOC (JPL)

2. Instrument Overview

The Scanning Habitable Environments with Raman & Luminescence for Organics & Chemicals (SHERLOC) instrument 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 co-registered 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.

SHERLOC's investigation combines two spectral phenomena, native fluorescence and pre-resonance/resonance DUV Raman scattering. These events are resolvable when a high-radiance, narrow line-width, laser source illuminates a sample. In fluorescence the incident photon is absorbed and re-emitted at a higher wavelength. The difference between the excitation wavelength and the emission wavelength is correlated with the number of electronic transitions between excitation and emission which increases with increasing aromatic structure (i.e., number of aromatic rings). Typical cross-sections are 10^5x greater than Raman, enabling the detection of sub-picograms of carbon.

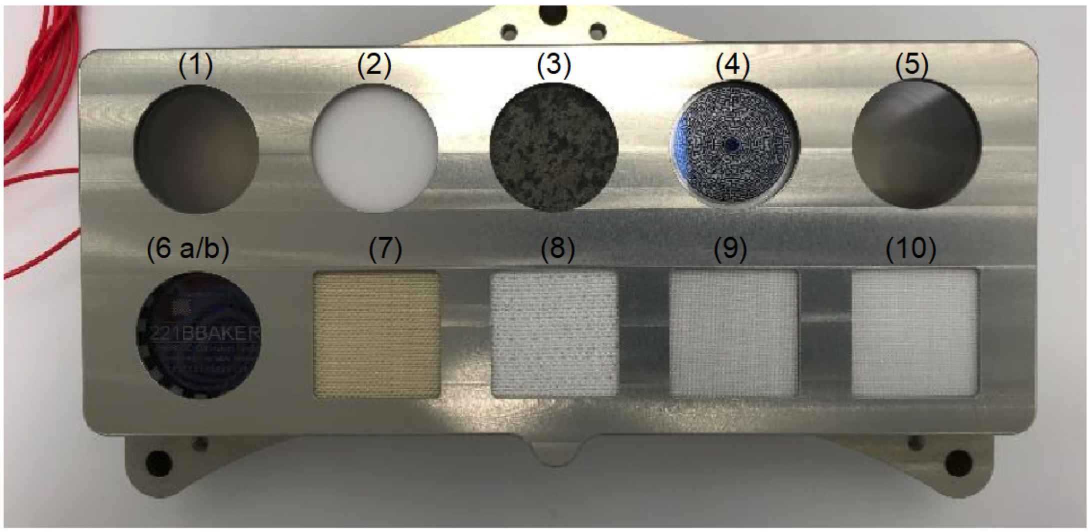
Native fluorescence emission of organics extends from ~270 nm into the visible. Conversely mineral fluorescence emission stemming from crystalline defects and impurities does not have strong absorption features in the deep UV resulting in mineral fluorescence that typically begins longward of 360 nm and continues through the visible into the NIR. The only reported fluorescence of naturally occurring inorganics in the region 250–360 nm is in non-relevant astrophysical conditions. In over 30 years of experiments in our laboratories, the only observed fluorescence at shorter wavelengths (<360 nm) have always been directly attributable to organics trapped inside the mineral matrix of a field sample, not the mineral matrix.

SHERLOC's narrow-linewidth (3 GHz) DUV laser (248.6 nm) also enables fluorescence-free Raman scattering for additional classification of aromatics and aliphatic organics and minerals. Raman scattering is inelastic scattering where the loss of energy from the excitation energy (248.6 nm) and defines the vibrational energy of a bond with which it interacted. This enables classification of bonds such as C-H, CN, C=O, C=C, NHx, NOx, SOx, POx, ClOx, and OH.

In addition to the two imaging subsystems and a DUV laser spectrometer, the Mars 2020 rover carries an external SHERLOC Calibration Target (SCT). The SCT is affixed to the front of the rover and includes ten calibration targets to enable instrument calibration on the Mars Surface.

Figure 1 captures the ten calibration targets and their location within the calibration target palette.

Figure 1 SHERLOC Calibration Target (SCT)



The calibration target material is being identified as:
(0) NA
(1) AlGaN on Sapphire Raman Region (~262 nm peak)
(2) Diffusil Diffuser (Bubbled Silica)
(3) Mars Meteorite SaU 008
(4) Intensity Target (Chromium on Silica)
(5) AlGaN on Sapphire Fluorescence Region (~335 nm peak)
(6) Polycarbonate Over Opal Glass Geocache
(7) Vectran Fabric
(8) Orthofabric
(9) Teflon Fabric
(10) nGimat Coated Teflon Fabric

For detail descriptive information on the SCT and each of the targets, refer to Ref [13], Ref [15], and Ref [18].

3. Data Processing Overview

3.1 Data Processing Level

This documentation uses the "Planetary Data System Standard 4" (PDS4) data level numbering system. See Table 2 for description of the PDS4 data processing levels.

The SHERLOC spectroscopy instrument operations data products referred to in this document as EDRs are considered processed at the "Raw" level and reconstructed from the "Telemetry" data stream, as defined by the Project Specific Formatted Data Unit (SFDU) record.

Spectroscopy RDRs are considered "Partially Processed", "Calibrated", or "Derived". The RDRs may be reconstructed from "Raw" data, or processed from other RDR data levels.

Table 2 Processing Level for Science Data Sets

PDS4	Operations Level	Description
Telemetry	n/a	An encoded byte stream used to transfer data from one or more instruments to temporary storage where the raw instrument data will be extracted. PDS does not archive telemetry data.
Raw	EDR	Original data from an instrument. If compression, reformatting, packetization, or other translation has been applied to facilitate data transmission or storage, those processes will be reversed so that the archived data are in a PDS approved archive format.
Partially Processed	EDR	Data that have been processed beyond the raw stage, but which have not yet reached calibrated status.
Partially Processed	RDR	Data that have been processed beyond the raw stage, but which have not yet reached calibrated status. These are more highly processed products than EDRs.
Calibrated	RDR	Data converted to physical units, which makes values independent of the instrument.
Derived	RDR	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.

3.2 Product Label Descriptions

There are two different sets of product labels (metadata) that can be associated with each RDR file, a) detached ODL label, b) detached PDS (version 4) label.

While label formats and keywords may differ, all can be reliably used to extract metadata. It is important to note that, with one exception, these labels contain the same semantic information and are in most cases in sync with each other. One exception is that the PDS4 label may contain items (such as comments, keywords, group and object delimiters) that are named differently than counterparts in the ODL labels, or may omit some items altogether.

The primary label supporting mission surface operations is the detached ODL label. ODL label is heritage from past Mars missions (e.g. MSL), and use a format similar to the now defunct PDS3 label standards. ODL labels are not included in the PDS archives.

The label from the archive perspective is the detached PDS4 label. This is a separate file with the same base filename as the RDR file, with a ".xml" extension.

3.2.1 PDS Labels

The NASA PDS adopted a new Planetary Data System Version 4 (PDS4) standard in 2013, which the Mars 2020 mission has agreed to use for data archiving. The SHERLOC archive data set will also be in compliance with the PDS4 standard. The archive data set labels for SHERLOC spectroscopy will be detached xml labels. This is a separate file with the same basename as the spectroscopy file, with an "xml" (lower case) extension. The detached PDS4 label is fully compliant with PDS4 archive standards, referenced in Section 1.3.

3.2.2 ODL Labels

As described in the Section 3.2, RDRs described in this document have a detached ODL label for operations and a detached PDS4 label for archiving. Each instrument is responsible for converting PDS-formatted products to be compatible with their own software systems.

The detached ODL label with its keyword = value format is very similar to deprecated PDS3 standards. A PDS3 label starts with the keyword assignment:

PDS_VERSION_ID = PDS3

The detached ODL label, which contains keywords that may not be compliant to PDS standards, starts with the keyword assignment:

ODL_VERSION_ID = ODL3

This first line distinguishes the two file types. With the exception of this keyword, any PDS3-format reader that does not validate against the PDS Data Dictionary should be able to read the ODL labels.

The ODL label is object-oriented, and describes the objects in the associated data file. The ODL label contains keywords for product identification, along with the data object definitions containing descriptive information needed to interpret or process the data in the file. The PDS3 SPREADSHEET object is used to describe the format of the products.

ODL label statements have the form of "keyword = value". Each label statement is terminated with a carriage return character (ASCII 13) and a line feed character (ASCII 10) sequence to allow the label to be read by many operating systems. Pointer statements with the following format are used to indicate the location of data objects in the file:

^object = location

where the carat character (^, also called a pointer) is followed by the name of the specific data object. The location is the 1-based starting record number for the data object within the file. Alternatively, it could be the 1-based byte location within the file if it includes a <bytes> unit tag.

Pointers are used to define the data object locations of the ASCII instrument data itself, e.g. ^COSMIC RAY CONFIGURATION PARAMETERS: = (SS__D004A0631370804_235RBR__0010256SRLC07305_0__C__J01.CSV,175<BYTES>)

3.2.2.1 Interpretation of N/A, UNK, and NULL

During the completion of data product labels or catalog files, one or more values may not be available for some set of required data elements. In this case PDS provides the symbolic literals “N/A”, “UNK”, and “NULL”, each of which is appropriate under different circumstances. As a note, if any one of these three symbolic literals are used in place of a keyword value that is normally followed by a Unit Tag(s) (e.g., “<*value*>”), the Unit Tag(s) is removed from the label.

“N/A” (“Not Applicable”) indicates that the values within the domain of this data element are not applicable in this instance. For example, a data set catalog file describing NAIF SPK kernels would contain the statement:

INSTRUMENT_ID = "N/A"

because this data set is not associated with a particular instrument.

“N/A” may be used as needed for data elements of any type (e.g., text, date, numeric, etc.).

“UNK” (“Unknown”) indicates that the value for the data element is not known and never will be. For example, in a data set comprising a series of images, each taken with a different filter, one of the labels might contain the statement:

FILTER_NAME = "UNK"

if the observing log recording the filter name was lost or destroyed and the name of the filter is not otherwise recoverable.

“UNK” may be used as needed for data elements of any type.

“NULL” is used to flag values that are *temporarily* unknown. It indicates that the data preparer recognizes that a specific value should be applied, but that the true value was not readily available. “NULL” is a placeholder. For example, the statement:

DATA_SET_RELEASE_DATE = "NULL"

might be used in a data set catalog file during the development and review process to indicate that the release date has not yet been determined.

“NULL” may be used as needed for data elements of any type.

Note that all “NULL” indicators should be replaced by their actual values prior to final archiving of the associated data.

4. Data Product Overview

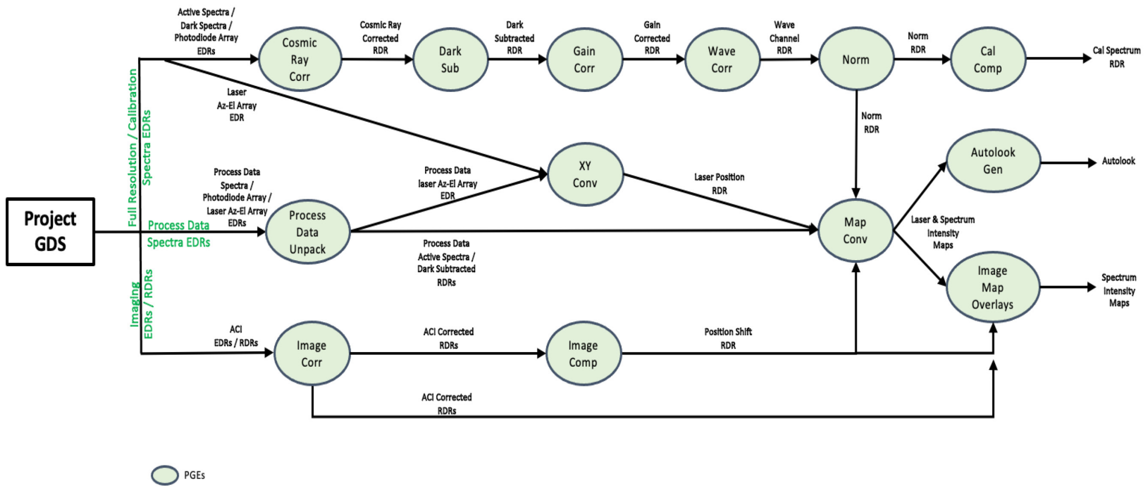
4.1 Data Flow

The SHERLOC spectroscopy EDR data products generated by the IDS element during surface operations is converted from the instrument raw telemetry data products. These spectroscopy EDRs are created on the M2020 Mission System (MS) Ground Data System (GDS) Amazon Web Services (AWS) environment, then deposited in the MS GDS Operations Cloud Store (OCS). The IDS EDR generation pipeline retrieves the pertinent raw telemetry to generate instrument EDRs.

For each SHERLOC spectroscopy EDR generated, a corresponding AWS Simple Notification system (SNS) message is created on the MS GDS. The SHERLOC instrument science data system (SDS) pipeline, operating outside of the MS GDS environment, intercepts the AWS SNS message. The SHERLOC SDS pipeline intercepts each AWS SNS and parses them to find the location of EDRs on GDS OCS. Thereafter the SDS pipeline retrieves the EDR and queues these EDRs within the SDS pipeline. Each queued EDR will wait for the invocation of pertinent science software algorithm executables or product generation executables (PGEs) to be called, then these EDRs are processed to generate the downstream RDRs. The SHERLOC SDS pipeline processes spectroscopy EDRs and ancillary instrument data to generate spatially resolved, spectrally-corrected spectrum intensity maps.

The SHERLOC SDS pipeline also calculates the checksum for each file that it ingests to detect file changes, and then determines whether to re-run a PGE and/or up-version the output data file. Therefore, RDRs are versioned, with the expectation that higher versioned files are superior to the lower versions of the same data file. Figure 2, below, depicts the SHERLOC SDS pipeline data workflow for the spectroscopy RDR data processing and product generations.

Figure 2 depicts the SHERLOC SDS Pipeline Data Workflow for the Spectroscopy RDR Data Product Generations



Below is a quick overview of the science EDRs that the SHERLOC SDS pipeline ingest:

- **Active Spectra (science target):**
  This type of EDR is marked with the ERA product ID in the filename.
  This product contains SHERLOC instrument SOH data and the active spectrum intensity values for each point measured in each detector region.

- **Dark Spectra (science target):**
  This type of EDR is marked with the ERB product ID in the filename.
  This product contains SHERLOC instrument SOH data and the dark spectrum intensity values for each point measured in each detector region.

- **Active Spectra (internal calibration target):**
  This type of EDR is marked with the ECA product ID in the filename.
  This product contains SHERLOC instrument SOH data and the active spectrum intensity values of the measured internal AlGaN calibration target for each point measured in each detector region.

- **Dark Spectra (internal calibration target):**
  This type of EDR is marked with the ECB product ID in the filename.
  This product contains SHERLOC instrument SOH data and the dark spectrum intensity values of the measured internal AlGaN calibration target for each point measured in each detector region.

- **Process_Data (science target):**

This type of EDR is marked with the ERP product ID in the filename.
This product contains SHERLOC instrument SOH data and the results of the internal Process_Data algorithm. The Process_Data algorithm may be commanded in a variety of ways, which will determine the data structure contained within this product. All products will contain a table of algorithms applied to the data. Spectral data (spectrally or spatially binned), scanner position data, laser photodiode data (all or averaged), and/or a ranked list of spectra may be also included.

The Process_Data command may command up to eleven algorithms for data processing, in any order. This will result in the generation of a variety of different data products, depending on how the data was processed, and what data is downlinked. Nominal Process_Data operations should obey the following rules:

- The first algorithm applied must be "Active Copy"
- "Zone Specification" must be applied some time after "Active Copy" or "Dark Subtraction" and before "Cosmic Ray Removal"
- "Ranked List" must be applied after all other algorithms (if at all), including "Zone Specification" and "Active Copy"
- "Laser Normalization" must be applied some time after "Active Copy" or "Dark Subtraction"
- "Laser Normalization" cannot be applied after "Spectral Binning", "Ranked List", or either of the "Photodiode Dump" algorithms
- "Bad Pixel Removal" must be applied some time after "Active Copy" or "Dark Subtraction"
- "Bad Pixel Removal" cannot be applied after "Spectral Binning", or "Ranked List", or either of the "Photodiode Dump" algorithms
- "Cosmic Ray Removal" must be applied some time after "Active Copy" or "Dark Subtraction"
- "Cosmic Ray Removal" cannot be applied after "Spectral Binning", or "Ranked List", or either of the "Photodiode Dump" algorithms
- "Spectral Binning" must be some time after "Active Copy", "Dark Subtraction" and, "Zone Specification"
- "Spatial Binning" must be applied some time after "Active Copy" or "Dark Subtraction"
- "Spatial Binning" cannot be applied after "Spectral Binning", "Ranked List", or either of the "Photodiode Dump" algorithms

A nominal algorithm order for process_data would be:

1. Active Copy
2. Dark Subtraction
3. Zone Specification
4. Cosmic Ray Removal
5. Laser Normalization
6. Spatial Binning

7. Raw Photodiode Data
8. Ranking

An overview of all Process_Data algorithms and the type of data that is expected to be produced:

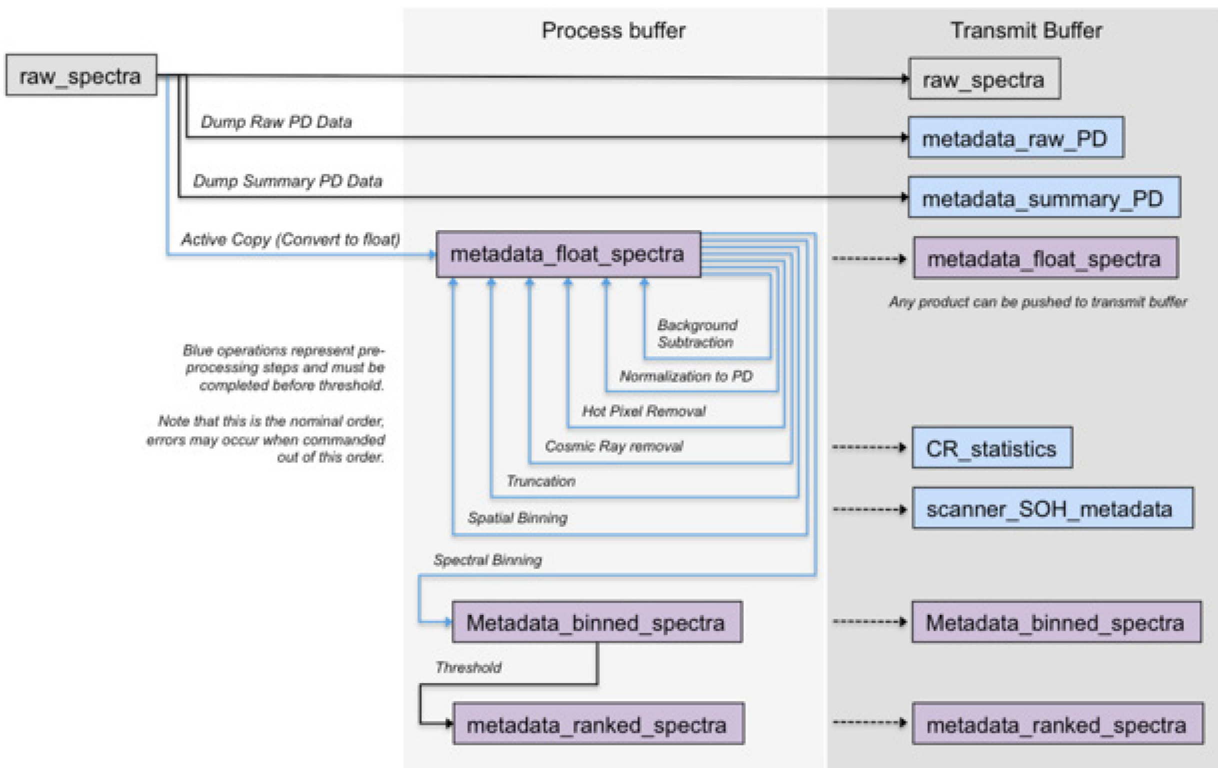
Table 3 Process_Data Algorithms

ID	Data Classification	Processing	Notes
N/A	Spectra	Active Copy	The active copy command will return the raw spectra stored in the data buffer. During nominal operations, this is always the first algorithm applied, and never the only algorithm.
1	Spectra	Background Subtraction	The dark frame is subtracted from the active frame. During nominal operations, this is always the second algorithm applied. The position table will be returned if this is the last command issued.
2	Spectra	Laser Normalization	The spectra (usually dark-subtracted, cosmic ray removed) is normalized to the photodiode laser intensity. During nominal operations, this command may be applied after cosmic ray-removal, or not at all.
3	Raw Photodiode Data	N/A	This algorithm will return all photodiode points.
4	Summary Photodiode Data	N/A	This algorithm will return the average photodiode value at each point.
5	Spectra	Bad Pixel Removal	This command will remove spectral intensity values associated with known bad pixels in the detector.
6	Spectra	Cosmic Ray Removal	Cosmic rays are removed from all spectra (usually dark-subtracted, not laser normalized). During nominal operations, this command is usually applied after background-subtraction but before laser normalization.
7	Spectra	Zone Specification	Up to six spectral regions (zones – start and stop locations) are specified to apply the cosmic ray removal, spectral binning, and ranking algorithms.
8	Binned Spectra Sum	Spectral Binning	This algorithm will reduce the number of spectral wavelength channels from 6444. The returned spectra will contain fewer spectral wavelength channels. During nominal operations, this algorithm is expected to be applied after all spectral processing steps
9	Spectra	Spatial Binning	A pre-defined spatial binning table is applied to the map to reduce the number of spectra from 1296 to 1-36. The returned spectra will include 1 to 36 spectra, rather than 1296 or 100 from the original map. During nominal

			operations, this algorithm is expected to be applied after all spectral processing steps
10	Ranked List	N/A	This algorithm produces a list of all spectra, ranked in order for follow-up scans

The following diagram shows all process_data algorithms, with blue arrows representing algorithms applied to spectral data and black arrows representing data pushed to the transmit buffer for GDS access.

Figure 3 Diagram of All Process_Data Algorithms



- **Laser Photodiode Data:**
  This type of EDR is marked with the EPA product ID in the filename.
  This product contains SHERLOC instrument SOH data and the laser photodiode data for each laser shot.

- **Scanner Position Data:**
  This type of EDR is marked with the ESP product ID in the filename.
  This product contains SHERLOC instrument SOH data and the commanded and retrieved scanner positions, as well as the scanner position error and the scanner current values.

4.2 File Naming

4.2.1 File Naming Standards

The file naming scheme has been relieved of the string length constraint imposed by the Level II 36.3 filename standard approved by PDS in 2009. PDS has migrated to a new standard that is PDS4, and filenames are essentially limited to a character string length of 255 characters. Use of three- character extensions, such as ".IMG" for image EDRs and RDRs and ".CSV" for spectroscopy EDRs, is consistent with the PDS standard.

The primary attributes of the filename nomenclature are:

a) Uniqueness - It must be unique unto itself without the file system's directory path. This protects against product overwrite as files are copied/moved within the file system and external to the file system, if managed correctly.

b) Metadata - It should be comprised of metadata fields that keep file bookkeeping and sorting intuitive to the human user. Even though autonomous file processing will be managed via databases, there will always be human-in-the-loop that puts a premium on filename intuition. Secondly, the metadata fields should be smartly selected based on their value to ground processing tools, as it is less CPU-intensive to extract information from the filename than from the label. NOTE: Metadata information in the filename also resides in the product label.

The metadata fields have been selected based on MER and PHX lessons learned. In general, the metadata fields are arranged to achieve:

a) Sortability - At the beginning of the filename resides a primary time oriented field such as Spacecraft Clock Start Count (SCLK). This allows for sorting of files on the MSL file system by spacecraft data acquisition time as events occurred on Mars.

b) Readability - An effort is made to alternate Integer fields with ASCII character fields to Optimize differentiation of field boundaries for the human user.

4.2.2 Spectroscopy RDR Filename

Each spectroscopy RDR filename uses the already established EDR file naming scheme. The file naming scheme's goal is to ensure each EDR/RDR is uniquely identified. The RDR data files share the same file naming scheme and convention as that of EDR data files to ensure consistency and conformity in both EDR and RDR. The established file naming scheme can be seen in Table 4.

Table 4 RDR File Naming Convention

1-2	3	4	5-8	9	10-19	20	21-23	24-26	27	28	29-31	32-35	36-44	45-51	52	53-54	55	56-58
INSTRUMENT	CONFIG	SPECIAL FLAG	SOL NUMBER	VENUE	SCLK	_	SUB-SCLK	PROD TYPE	GEOMETRY	THUMBNAIL	SITE	DRIVE	SEQUENCE/RTT	INST. SPECIFIC	PRODUCER	VERSION	.	EXT

Table 5 below, provides additional detail into the spectroscopy RDR file naming field and how each field is used in defining the SHERLOC spectroscopy RDR filename.

Table 5 RDR File Naming Fields

Field Name	Field Position (Size)	Data Type	Description
instrument	01 (2)	ASCII	Instrument Identifier. The valid value is: SS : denotes SHERLOC Spectroscopy subsystem Other applicable identifiers for the SHERLOC imaging subsystems are (but not used in this SIS): SI : denotes SHERLOC WATSON subsystem SC : denotes SHERLOC ACI subsystem
config	03 (1)	ASCII	Instrument Configuration, an operational attribute of the instrument that assists in characterizing the data: Not used and set to "_".
special flag	04 (1)	ASCII	Special Processing flag, applicable to RDR only. Not used and set to "_".
sol number	05 (4)	ASCII	Solar Day Number, corresponding to the time conversion of SCLK + Sub-SCLK time when spectroscopy data was acquired.
venue	09 (1)		Venue type denoting the data processing context and activity. The valid values are: "_" : Flight (surface or cruise)
sclk	10 (10)	Integer	Spacecraft Clock in second time unit, corresponding to the time the spectroscopy data was acquired by the instrument.

<table>
<tr><td>_</td><td>20 (1)</td><td>ASCII</td><td>Separator for SCLK and Sub-SCLK field. Always set to “_”</td></tr>
<tr><td>sub-sclk</td><td>21 (3)</td><td>Integer</td><td>Spacecraft Clock in sub-second time unit, corresponding to the time the spectroscopy data was acquired.</td></tr>
<tr><td>product id</td><td>24 (3)</td><td>ASCII</td><td>Product Identifier. See Section 4.4 for the Product ID description.</td></tr>
<tr><td>geometry</td><td>27 (1)</td><td>ASCII</td><td>Not used and set to “_”.</td></tr>
<tr><td>thumbnail</td><td>28 (1)</td><td>ASCII</td><td>Not used for SS products and set to “_”.</td></tr>
<tr><td>site</td><td>29 (3)</td><td>Integer</td><td>Site identifier (see Section 5.3.4 on Site frame). Site location count from the RMC where the data was acquired.
<table>
<tr><th>Values</th><th>Range</th></tr>
<tr><td>000, 001, …, 999</td><td>0 thru 999</td></tr>
<tr><td>A00, A01, …, A99</td><td>1000 thru 1099</td></tr>
<tr><td>B00, B01, …, B99</td><td>1100 thru 1199</td></tr>
<tr><td>…</td><td>…</td></tr>
<tr><td>Z00, Z01, … Z99</td><td>3500 thru 3599</td></tr>
<tr><td>AA0, AA1, …, AA9</td><td>3600 thru 3609</td></tr>
<tr><td>AB0, AB1, …, AB9</td><td>3610 thru 3619</td></tr>
<tr><td>…</td><td>…</td></tr>
<tr><td>ZZ0, ZZ1, …, ZZ9</td><td>10350 thru 10359</td></tr>
<tr><td>AAA, AAB, …, AAZ</td><td>10360 thru 10385</td></tr>
<tr><td>ABA, ABB, …, ABZ</td><td>10386 thru 10411</td></tr>
<tr><td>…</td><td>…</td></tr>
<tr><td>ZZA, ZZB, …, ZZZ</td><td>27910 thru 27935</td></tr>
<tr><td>0AA, 0AB, …, 0AZ</td><td>27936 thru 27961</td></tr>
<tr><td>0BA, 0BB, …, 0BZ</td><td>27962 thru 27987</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><td></td><td></td><td></td><td><table>
<tr><td>…</td><td>…</td></tr>
<tr><td>7CA, 7CB, …, 7CZ</td><td>32720 thru 32745</td></tr>
<tr><td>7DA, 7DB, …, 7DV</td><td>32746 thru 32767</td></tr>
<tr><td>_ _ _</td><td>Value out of range</td></tr>
</table></td></tr>
<tr><td>drive</td><td>32 (4)</td><td>Integer</td><td>Drive identifier. Drive count (position within a Site location) from RMC where the data was acquired.
<table>
<tr><td>Values</td><td>Range</td></tr>
<tr><td>0000, 0001, …, 9999</td><td>0 thru 9999</td></tr>
<tr><td>A000, A001, …, A999</td><td>10000 thru 10999</td></tr>
<tr><td>B000, B001, …, B999</td><td>11000 thru 11999</td></tr>
<tr><td>…</td><td>…</td></tr>
<tr><td>Z000, Z001, … Z999</td><td>35000 thru 35999</td></tr>
<tr><td>AA00, AA01, …, AA99</td><td>36000 thru 36099</td></tr>
<tr><td>AB00, AB01, …, AB99</td><td>36100 thru 36199</td></tr>
<tr><td>…</td><td>…</td></tr>
<tr><td>AZ00, AZ01, …, AZ99</td><td>38500 thru 38599</td></tr>
<tr><td>BA00, BA01, …, BA99</td><td>38600 thru 38699</td></tr>
<tr><td>BB00, BB01, …, BB99</td><td>38700 thru 38799</td></tr>
<tr><td>…</td><td>…</td></tr>
<tr><td>LJ00, LJ01, …, LJ35</td><td>65500 thru 65535</td></tr>
<tr><td>_ _ _ _</td><td>Value is out of range</td></tr>
</table></td></tr>
<tr><td>seq-id</td><td>36 (9)</td><td>ASCII</td><td>Instrument sequence-ID, the command sequenced of the spectroscopy or imaging measurement to be acquired. The first four characters are always SRLC. The remaining five characters are digits. The first digit after SRLC will always be 1 for</td></tr>
</table>

				spectral observations, and 0 for WATSON observations.
inst specific	1st proc flag	45 (1)	ASCII	Instrument specific identifier. The valid values are: "W" : denotes Wavelength Correction algorithm performed "B" : denotes that the on-board Process_Data product is background-subtracted "_" : denotes no Wavelength Correction algorithm performed
	exp id	46 (1)	ASCII	Instrument specific identifier to denote spectroscopy experiment ID designation. Values / Range 0, 1, …, 9 / 0 thru 9 A, B, …, Z / 10 thru 36
	img num	47 (2)	ASCII	Instrument specific identifier. The valid value is the total number of ACI images acquired, in relation to the ExpID. The ACI image number can range between "00" thru "99".
	2nd proc flag	49 (1)	ASCII	Instrument specific identifier. The valid values are: "C" : denotes Cosmic Ray Correction algorithm performed "Z" : denotes that Cosmic Ray Correction has been skipped, according to the algorithm configuration file "_" : denotes no Cosmic Ray Correction algorithm performed
	3rd proc flag	50 (1)	ASCII	Instrument specific identifier. The valid values are: "G" : denotes Gain Correction algorithm performed "P" : denotes that the on-board Process_Data algorithm has been applied (no gain correction is ever applied to Process_Data algorithms) "_" : denote no Gain Correction algorithm performed

<table>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>4th proc flag</td><td>51 (1)</td><td>ASCII</td><td>Instrument specific identifier. The valid values are:
“N” : denotes the laser normalization algorithm performed
“Z” : denotes that laser laser normalization has been skipped, according to the algorithm configuration file
“_” : denotes no laser normalization Correction algorithm performed</td></tr>
<tr><td colspan="2">producer</td><td>52 (1)</td><td>ASCII</td><td>Identifier for the institution/team that created this product.
“J” : JPL (SHERLOC PI Institution)</td></tr>
<tr><td colspan="2">version</td><td>53 (2)</td><td>ASCII</td><td>Product version number. The valid values are:
<table><tr><th>Values</th><th>Range</th></tr><tr><td>01, 02 …, 99</td><td>1 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></table></td></tr>
<tr><td colspan="2">.</td><td>55 (1)</td><td>ASCII</td><td>Separator for filename and file extension. Always set to “.”</td></tr>
<tr><td colspan="2">ext</td><td>56 (3)</td><td>ASCII</td><td>File extension. The valid value is:
“csv” : comma separate value file
“xml” : detached PDS4 label file
“LBL” : detached Ops ODL label file
“TIF” : TIFF formatted image file (no label)</td></tr>
</table>

4.3 Data Storage Convention

The SHERLOC Spectroscopy RDR data files are stored as ASCII tables and in comma separate value (CSV) format. Each RDR data structure comprises multiple tables. Tables with the same header prefix, e.g. Cosmic_Ray_List, will have the same number of table columns.

The detached ODL labels accompanying Spectroscopy RDRs for operations are stored as ASCII text, with each keyword definition terminated by ASCII carriage-return and line-feed characters. The ODL labels are not to be archived.

A separate PDS4 standard archive label will be generated during archive processing prior to periodic PDS archive delivery.

4.4 RDR Data Products Descriptions

Table 6 provides a quick view of the parent EDR to first-order child RDR relationships. The instrument SDS is responsible for generating tactical RDRs, generating PDS-compliant detached PDS4 labels, validating the labels, and performing PDS bundle collection for the spectroscopy RDRs prior to formal delivery of the bundle collection to PDS.

Table 6 Relationships between Spectroscopy EDRs and 1st Order RDRs

EDR Product Type Description	Parent EDR Product ID	Child RDR Type Description	1st Order RDR Product ID
SOH in Calibration Mode	"ECH"	n/a	n/a
SOH in Science Mode	"ESH"	n/a	n/a
Scanner Position Array (in millimeter)	"ESP"	Scanner Position Array (in image pixel position)	"RLP"
Photodiode Array	"EPA"	n/a	n/a
Full Resolution Active Spectra in Calibration Mode	"ECA"	Cosmic Ray Corrected Active Spectra & Cosmic Ray Identification list in Calibration Mode	"RCA" & "RAC"
Full Resolution Dark Spectra in Calibration Mode	"ECB"	Cosmic Ray Corrected Dark Spectra & Cosmic Ray Identification list in Calibration Mode	"RCB" & "RBC"
Full Resolution Active Spectra in Science Mode	"ERA"	Cosmic Ray Corrected Active Spectra & Cosmic Ray Identification list in Science Mode	"RRA" & "RAR"
Full Resolution Dark Spectra in Science Mode	"ERB"	Cosmic Ray Corrected Dark Spectra & Cosmic Ray Identification list in Science Mode	"RRB" & "RBR"

Process_Data Spectra	"ERP"	Processed Active Spectra & Dark-Subtracted Spectra	"RRA" & "RRS"
Processed Photodiode Array	"EPA"	n/a	n/a
Processed Scanner Position Array	"ESP"	Scanner Position Array (in image pixel position)	"RLP"

Tactically generated RDRs will be archived, other than Autolooks. The following sections provide more detailed descriptions of RDR product types, in reference to the Figure 2.

4.4.1 Cosmic Ray Correction RDRs

Prod ID: RRA, RRB, RAR, RBR and/or RCA, RCB, RAC, RBC

Cosmic rays are identified in each active and dark frame spectrum. Each group of channels (2-4 pixels) of the active or dark frame spectrum associated with a cosmic ray is replaced with an average or interpolation between spectral intensity points outside of the cosmic ray zone after dark subtraction in the next step described below (Dark Subtraction). Cosmic rays candidates are selected and replaced using the method described in Ref [14].

Below is an overview of the cosmic ray candidate selection steps:

1. A histogram is generated for all wavelength channels (intensity values are binned)
2. A cosmic ray is detected when a bin lies beyond a configurable threshold value, and at least one empty bin precedes the cosmic ray candidate.
3. The active and dark spectra and a list of cosmic ray candidate locations are generated as output products.

Active and dark frame RDRs are stored as several tables in the CSV file of size N x 2148 x R, where N is the number of spectrum points and R is the number of regions, represented by tables in the CSV file. The SCCD is usually configured to readout 3 detector regions of variable gain (but 1-5 regions are possible), and 2148 is the number of wavelength channels, represented by columns in each table. Each table column size could vary in size as one ASCII_STRING (one byte), one ASCII_REAL (8 bytes), or ASCII_INTEGER (4 bytes).

The list of cosmic ray candidate locations RDR is stored as multiple tables in the CSV file. The first table object stores the cosmic ray removal configuration parameters where each column represents the following parameter information:

- replacement_method: The method to use for replacing cosmic rays: I for interpolation between points or A for average values between points
- threshold_multiplier: The threshold multiplier to apply to histogram
- sigma_clip_val: sigma clipping threshold value
- channel_low: skip the first N pixels (usually set to ignore the inactive SCCD region)

- channel_high: skip the last N pixels (usually set to ignore the inactive SCCD region)
- cosmic_ray_limit: the limit to the number of cosmic rays that can exist in a spectrum – if this is exceeded, the spectrum probably has a high baseline causing false positive detections
  histogram_limit: the minimum number of map points necessary to apply histogram removal algorithm. If the number of spectra is below this value, the 1st derivative method is used.
- threshold_der: the first derivative threshold limit, if the first derivative method is used

The technique used in this PGE to identify cosmic rays is described in Ref [14]. Essentially, a histogram is generated for all wavelength channels - intensity values are binned. A cosmic ray is detected when a non-zero bin lies past the threshold value and at least one empty bin precedes the cosmic ray candidate. Cosmic rays may be replaced by one of two functions: 1) the cosmic ray is replaced by an average of the surrounding values, or 2) the cosmic ray is replaced by a line of best fit to the surrounding points. This method may fail in noisy regions of spectra. If the algorithm fails to interpolate surrounding points, the average replacement method will be used.

The subsequent tables, depending on the number of spectra region measurements taken, represent spectra regions. Each region table is size of N x 3 where N represents the number of cosmic rays identified (represented as rows in the CSV file). The first column defines the spectrum number (integer), the second column defines the detector channel that the cosmic ray was identified in (integer), and the third column defines the spectrum intensity of the cosmic ray.

If an insufficient number of spectral mapping points are available (usually less than 20), a first derivative method is used instead. In this case, the first derivative is calculated for each spectrum, and any points above a configurable threshold is flagged as a cosmic ray noise spike. This method generates more false positives than the histogram replacement method described in Ref [14], and the cosmic ray configuration file may require wavelength selection ranges to avoid the Raman region to prevent smoothing Raman peaks. Under nominal activity modes, the first derivative cosmic ray removal method is not expected to be implemented.

4.4.2 Dark Subtraction RDRs

Prod ID: RRS and/or RCS with special processing flag “C” in the filename
The SHERLOC dark frame is subtracted from the active frame for each SCCD region. Dark frame spectra are collected during each spectral measurement at each map position by exposing the SCCD for the same duration as an “active” exposure, but suppressing the laser trigger. Dark frame subtraction corrects for dark current, pixel-to-pixel variability, and hot pixels.

The cosmic ray candidate list generated in the Cosmic Ray Correction PGE is used to remove cosmic rays from the dark-subtracted spectra using the interpolation or average replacement method described above.

The dark subtracted, cosmic-ray removed RDR is stored as multiple tables in the CSV file (one for each detector region) of size N x 2148, where N is the number of spectrum points, and 2148 represents the number of wavelength channels (represented by columns in each table).

4.4.3 Gain Correction RDRs

Prod ID: RRS and/or RCS with special processing flag "G" in the filename
The SCCD gain may be defined independently for each region to provide higher CCD sensitivity in Raman spectral regions than fluorescence spectral regions, where a higher fluorescence signal is expected. Gain correction allows for direct comparison between each spectral region.

The gain correction RDR is stored as multiple tables in the CSV file (one for each detector region) of size N x 2148, where N is the number of spectrum points, and 2148 represents the number of wavelength channels (represented by columns in each table).

4.4.4 Wavelength Correction RDRs

Prod ID: RRS and/or RCS with special processing flag "W" in the filename
Raman peak locations of the internal AlGaN target and the suite of calibration targets were used to empirically define a polynomial relationship between SCCD channel number and wavelength prior to landing (see Ref [18]). Polynomial coefficients are updated during routine calibration measurements to correct for potential instrument alignment changes. A table of wavelength values is appended to each spectral RDR product.

The wavelength correction RDR is stored as a single table in the CSV file of size 1 x 2148. The gain-corrected spectra are also carried over in the RDR product.

4.4.5 Normalization RDRs

Prod ID: RRS and/or RCS with special processing flag "N" in the filename
The intensity of the SHERLOC laser varies over the duration of a measurement primarily due to temperature fluctuations in the laser power supply. The laser photodiode response is a function of the energy deposited on the target by the laser; spectra must be normalized to the maximum laser energy to account for a potentially variable source to directly compare spectra within a measurement.

The laser intensity normalization RDR is stored as multiple tables in the CSV file (one for each detector region) of size N x 2148, where N is the number of spectrum points, and 2148 represents the number of wavelength channels (represented by columns in each table). The wavelength calibration is included in the RDR as well.

4.4.6 Calibration Comparison RDRs

Prod ID: RCC

SHERLOC measurements of the internal AlGaN target are routinely collected to characterize the performance of the instrument over time. The internal AIGaN target is located inside the instrument dust cover.

Fermi-Dirac and Gaussian functions are fit to the most prominent Raman features in the spectra, as well as the laser reflection line, to determine the peak characteristics of each measurement. The peak position and shape of all features in each spectrum are trended to evaluate the performance and alignment of the SHERLOC spectrometer over the lifetime of the mission to recalibrate the wavelength calibration when necessary.

The Calibration Comparison RDR is stored as single table in the CSV file with 7 columns, where column denotes the Sol, SCLK, laser reflection peak location, laser reflection FWHM, AlGaN peak location, and AlGaN FWHM.

4.4.7 Image Correction RDR

Prod ID: RLC

Raw ACI and WATSON images are radiometrically and geometrically corrected by the MS IDS pipeline. The SHERLOC SDS pipeline may not need to perform post-processing of these IDS-generated image RDR(s), but will perform a ‘pass-thru’ of these RDRs to the subsequent SDS pipeline PGE for downstream data processing. The Image Correction RDR will not be archived.

4.4.8 Image Comparison RDR

Prod ID: RLD

Before, after, and throughout a nominal SHERLOC spectral measurement, ACI images of the target are acquired. A feature recognition algorithm is applied to all ACI images collected as part of an observation, images are aligned, and offsets between images are calculated. Under ideal laboratory conditions, no offset would exist between images. During operations on Mars, slight offsets are expected between images due to environmental effects (wind, thermal expansion of mechanical components caused by temperature variability) and robotic arm drift.

These offsets are characterized by image transformation parameters, capturing the linear offset, rotation, and varying distance to the target. The image transformation RDR contains a single table in a CSV file with the raw image file name, the corrected image file name, and the associated linear shift of the image, relative to the first image in the series. The RDR will compose of 6 columns: orig_image_name, image_name, x, y, z, rotation.

The Image Comparison RDR will not be archived. The results of this processing are carried over to subsequent PGEs, and this data is specifically stored in the RLS and RMO RDRs.

4.4.9 XY Conversion RDR

Prod ID: RLP

This product lists the spectroscopy scanner position of each laser shot in millimeters (mm) in the target space.

The XY Conversion RDR will not be archived. The results of this processing are carried over to subsequent PGEs, and this data is specifically stored in the RLS and RMO RDRs.

4.4.10 Map Conversion RDRs

Prod ID: RMx, RLI and RLS

The integrated intensity within a specified wavelength region for each point is calculated and positioned on a greyscale map based on the drift-corrected laser shot positions in the target plane. Each of the (up to) 35 wavelength ranges represents a Raman or fluorescence peak or region of interest that may be modified or configured to better suit a specific target.

Several RDRs are generated as part of the map conversion algorithm:
RM[0-9,A-Z (not O)]: A single column table with the integrated intensity map values for each point in a scan.
RLI: A single column table with the laser intensity map values for each point in a scan.
RLS: Drift-corrected laser shot positions for each ACI image. This RDR contains multiple tables (one for each ACI image acquired) with five columns: an index number, the spectrum name, the image name, and the pixel positions on the image of each laser shot

Each Map Conversion RDR is stored as single table in the CSV file with various number of columns.

4.4.11 Image Map Overlays RDRs

Prod ID: RMO

This product summarizes spectral intensity, position, and imaging data, stored in several tables in a CSV file. The first table contains the drift-corrected laser shot position and the image associated with these positions. The second table contains the wavelength range for the spectral intensity value. The third table contains the spectral intensity values. The column_index and position_index values may be used to link each table.

LASER_SHOT_POSITIONS: contains the image RDR file name and position of each laser shot on that image
Columns: Image_name, Position_index, x, y

WAVELENGTH_REGIONS: The wavelength range (start and stop wavelength values) for each spectral intensity value listed in the SPECTRAL_INTENSITY table. The column index pairs the wavelength with the column number (starting at index 0) in the SPECTRAL_INTENSITY table. For a discrete wavelength point, the wavelength_start and wavelength_stop values will be identical.
Columns: Column_index, Wavelength_start, Wavelength_stop

SPECTRAL_INTENSITY: Up to 35 columns (one column for each WAVELENGTH_REGIONS row) containing the spectral_intensity of the associated wavelength region. The Position_index value is matched with the Position_index value in the LASER_SHOT_POSITIONS table to associate spectral intensity values with positions in the image.
Columns: Position_index, Spectral_intensity_[0-34]

4.4.12 Process_Data RDRs

Prod ID: RRA or RRS special processing flag "P" in the filename
The process_data unpack algorithm characterizes and further processes (if necessary) the process_data EDR. The algorithm determines if cosmic ray correction should be applied to the data set, applies gain correction, if necessary, applies laser normalization, if necessary, and outputs the wavelength-calibration spectrum intensity values in a CSV.

The resulting RDR is stored as multiple tables in the CSV file (one for each detector region) of size N x 2148, where N is the number of spectrum points. 2148 represents the number of wavelength channels (represented by columns in each table). The wavelength calibration is included in the RDR as well.

4.4.13 Autolook

Prod ID: SAL
The Spectroscopy Autolook report is a project directed science-oriented decisional tactical data file. This type of data files requires no human-in-the-loop processing due to the quick turnaround time requirement of 15-20 minutes from the moment the SHERLOC EDRs are available on the GDS OCS. This report depends on many science RDRs, both spectroscopy and imaging, as input files to trigger the execution and generation. A final report format in both PDF and PNG will be uploaded to GDS OCS and are available for Project users. The Autolook file naming and data format convention conforms to the Mars 2020 Autolook and Quicklook Data Product Software Interface Specification [JPL D-105237].

The SHERLOC Spectroscopy Autolook example filename is as follows:

ASRLC_SOL0200_SAL_0625428141_015_0010044_SRLC07006_W0__CGN_J01.PDF

Autolooks will not be archived at the PDS.

The description provided below is provided as an added reference to help users understand the content and purpose of this product. Spectroscopy Autolook products will be generated from the following input science RDR files:

- Raw, unprocessed dark spectrum data (science or calibration targets) (ERB, ECB product IDs)
- Processed spectral data (science or calibration targets) (RRS, RCS product IDs) with the following processing applied:
  - Cosmic ray correction
  - Dark subtraction
  - Gain correction
  - Wavelength correction
  - Laser photodiode normalization
- Process_data products (RRA or RRS products)
- The following additional products, if available:
  - Spectral intensity maps (RMx product IDs)
  - Laser photodiode intensity values (EPA product ID)
  - Scanner position values (RLP product ID)
  - Laser shot locations (RLS product ID)
  - WATSON image
  - ACI image

Each data product will be displayed on a pdf (two pages long), using the layout provided in the example below. For single-spectra products, only the first page of the Autolook product will be generated.

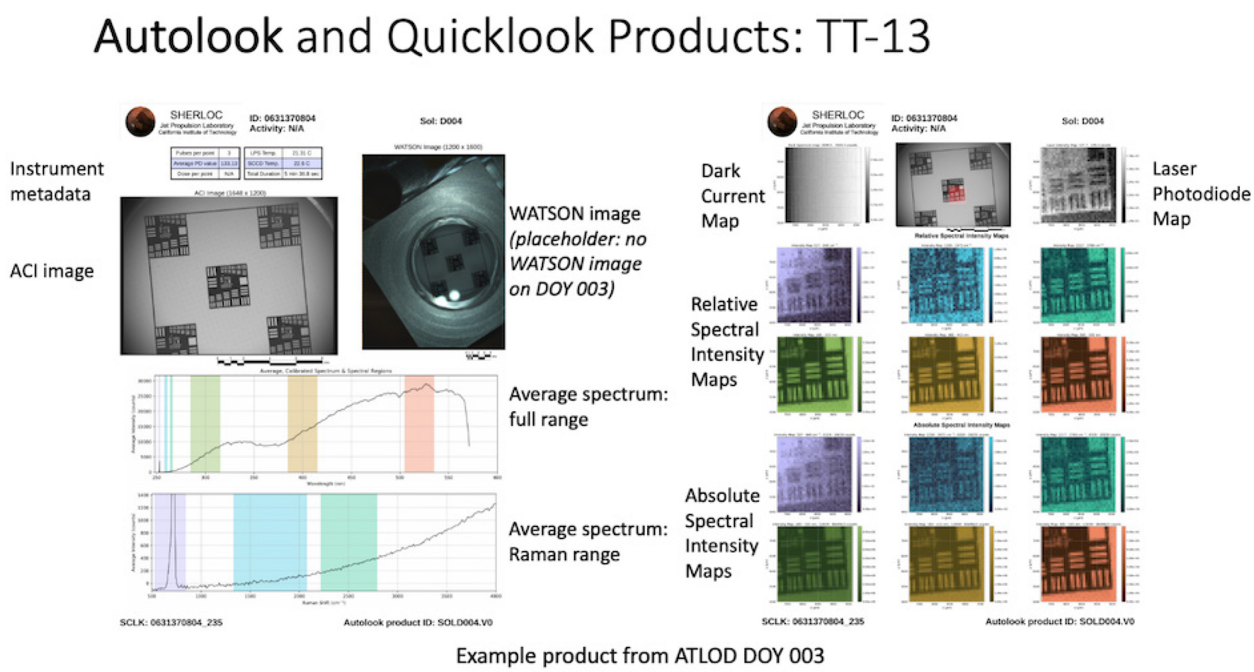
The Autolook header contains information on the target, sequence ID, and the Sol. A metadata table contains the average photodiode value, the number of pulses per point, the total integration time, and the temperature of the SCCD and the laser power supply.

The first ACI and WATSON images are displayed, with scale bars. An average spectrum of the entire wavelength region, and the Raman region are also displayed.

The second page contains an annotated ACI image, showing each of the laser spot locations, a map of the average dark spectrum at each point, and a map of the average laser photodiode value at each point.

Two versions of the spectrum intensity maps are generated: the first set scales of each map within the spectral range of an individual spectral region (relative spectral intensity maps). The second set of maps scales the first three maps (Raman region) and the second three maps (fluorescence region) together (absolute intensity maps). The color of each spectral intensity map corresponds to the shaded region in the average spectra on page one of the Autolook report. These spectral regions can be modified by changing the configuration file in the SDS pipeline.

Figure 4 Autolook Example



4.5 Data Product Format

As stated in Section 4.4, SHERLOC Spectroscopy RDR science data files are ASCII formatted CSV files, with the exception of Autolook products. A detached ASCII ODL label will accompany each RDR data file, during the data processing. A separate PDS4 label will also be generated with each RDR data file when assembling archive data set for PDS delivery.

The ASCII-formatted RDR data file will have a ".CSV" extension, while the accompanying detached PDS4 label will have a ".xml" extension.

4.6 Product Validation

SHERLOC PGEs are tested on data sets collected during ATLO and I&T instrument development and testing phases to ensure that data is processed correctly. Unit testing is conducted on each PGE and methods shared by PGEs to validate the expected outcome of each algorithm. Test data sets include measurements of each calibration target (on the calibration target palette and the internal AlGaN target) collected during the system thermal test, under similar environmental conditions as those expected on Mars.

Validation of products for PDS compliance are also conducted using the NASA PDS `validate` tool (https://nasa-pds.github.io/validate/) and compared against the PDS-approved Mars 2020 data dictionary (https://pds. nasa.gov/datastandards/dictionaries/). The `validate` tool will be used to ensure that the labels generated by each PGE follow the appropriate schema, and tables saved in CSV RDR files are checked to ensure that they are formatted correctly, column headers match expected values defined in labels, and values match the defined data type and exist within the expected range. Individual RDR products are validated as they are designed, and PDS bundles are validated prior to delivery.

5. Standard Used in Generating Products

5.1 PDS Standards

The instrument RDR data product complies with Planetary Data System standards for file formats and labels, as specified in the PDS4 Standards Reference in Section 1.3.

5.2 Time Standards

The RDR PDS label uses attributes containing time values. Each time value standard is defined according to the attribute description. Please refer to SHERLOC Spectroscopy EDR SIS Appendix B.

5.3 Coordinate Frame Standards

The M2020 Frame Manager defines several dozen coordinate frames, which can be used for command pointing among other things. Refer to the Pointing, Positioning, Phasing and Coordinate Systems (PPPCS) document in Ref [7] or the Surface Attitude, Positioning and Pointing (SAPP) Functional Design Description (FDD) in Ref [8] for more details on all these coordinate frames. Only a few of them are used by the products and processes described by this SIS. This subset is described in detail in this section. The only place in this SIS where the full set of frames can appear is in the INSTRUMENT_COORD_FRAME_ID label, which is a command echo.

A subset of these frames needed for a specific image or data set are defined by the *_COORDINATE_SYSTEM groups in the ODL label.

Table 7 Coordinate Frames Used for M2020 Surface Operations

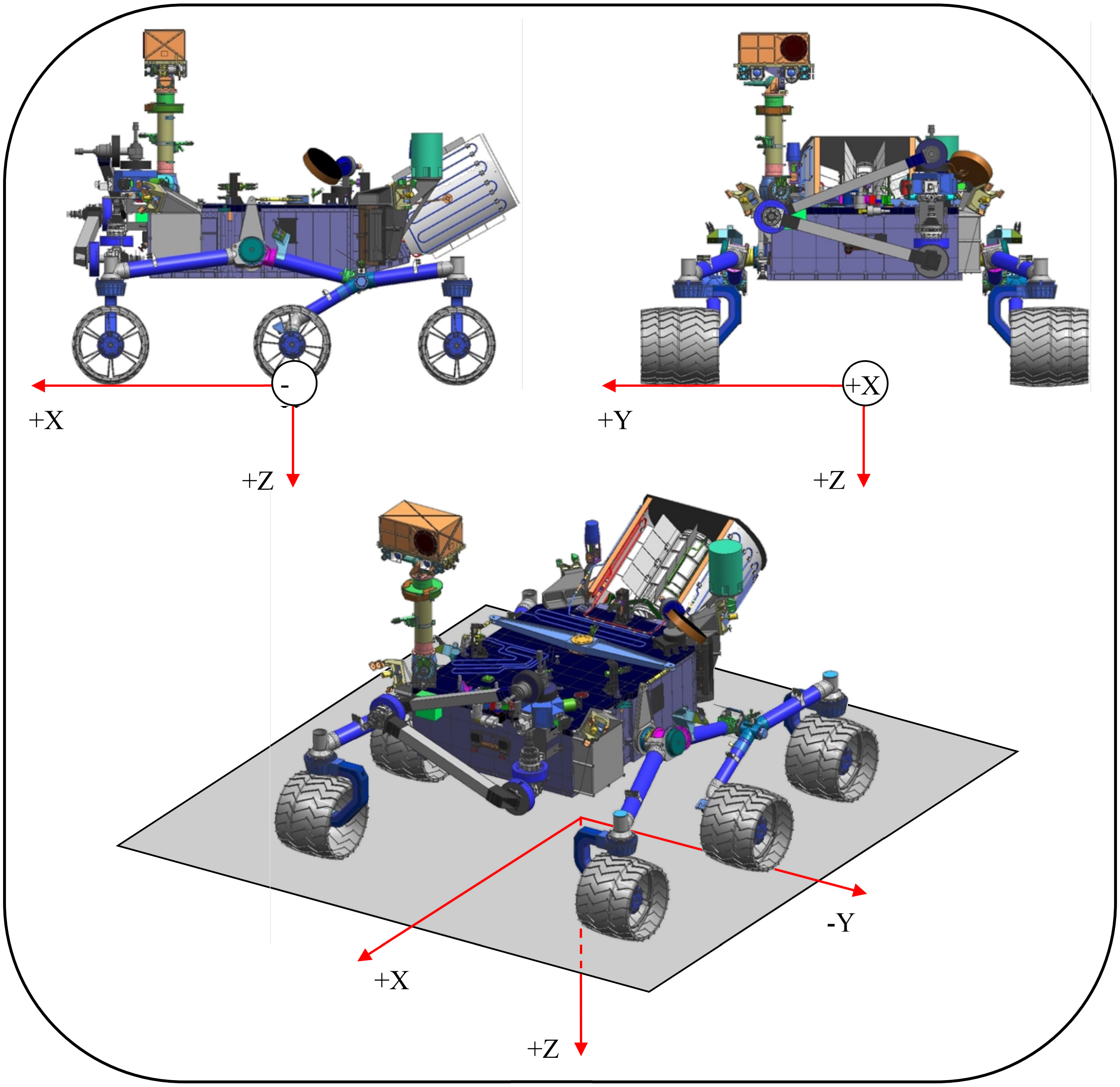
Frame Name (Label Keyword Value)	Short Name (SAPP FDD)	Reference Frame (Used to Define)	Coordinate Frame	
			Origin	Orientation
ROVER_NAV_FRAME	RNAV	Enclosing SITE_FRAME	Attached to rover	Aligned with rover
ROVER_MECH_FRAME	RMECH	Enclosing SITE_FRAME	Attached to rover	Aligned with rover
LOCAL_LEVEL_FRAME	LL	Enclosing SITE_FRAME	Attached to rover (coincident with Rover Nav Frame)	North/East/Nadir
SITE_FRAME	SITE(n)	Previous SITE_FRAME	Attached to surface	North/East/Nadir
RSM_HEAD_FRAME	RSM_HEAD	ROVER_NAV_FRAME	Attached to mast head	Aligned with pointing of mast head. This corresponds to RSM_HEAD in the Frame Manager

Frame Name (Label Keyword Value)	Short Name (SAPP FDD)	Reference Frame (Used to Define)	Coordinate Frame	
			Origin	Orientation
Arm Frames: ARM_BASE_FRAME ARM_TURRET_FRAME ARM_DRILL_FRAME ARM_GDST_ FRAME ARM_TOOL_FRAME ARM_WATSON_FRAME ARM_SHERLOC_FRAME ARM_PIXL_FRAME	Arm Frames: RA_BASE TURRET DRILL GDRT TOOL WATSON SHERLOC PIXL	ROVER_NAV_FRAME	Attached to the tool; see PPPCS for the specific tool frame.	Aligned with tool in some way; see PPPCS for the specific tool Frame.

5.3.1 Rover Navigation (Rover Nav) Frame

The Rover Nav frame (RNAV) is the reference frame used for surface navigation and mobility. By definition, the frame is attached to the rover, and moves with it when the rover moves while on the surface. Its Y origin is centered on the rover and the X origin is aligned with the middle wheels' rotation axis for the deployed rover and suspension system on a flat plane. The Z origin is defined to be at the nominal surface, which is a fixed position with respect to the rover body. The actual surface will likely not be at exactly Z=0 due to the effects of suspension sag, rover tilt, rocker bogie angles, etc. The +X axis points to the front of the rover, +Y to the right side, and +Z down (perpendicular to the chassis deck). See Figure 5.

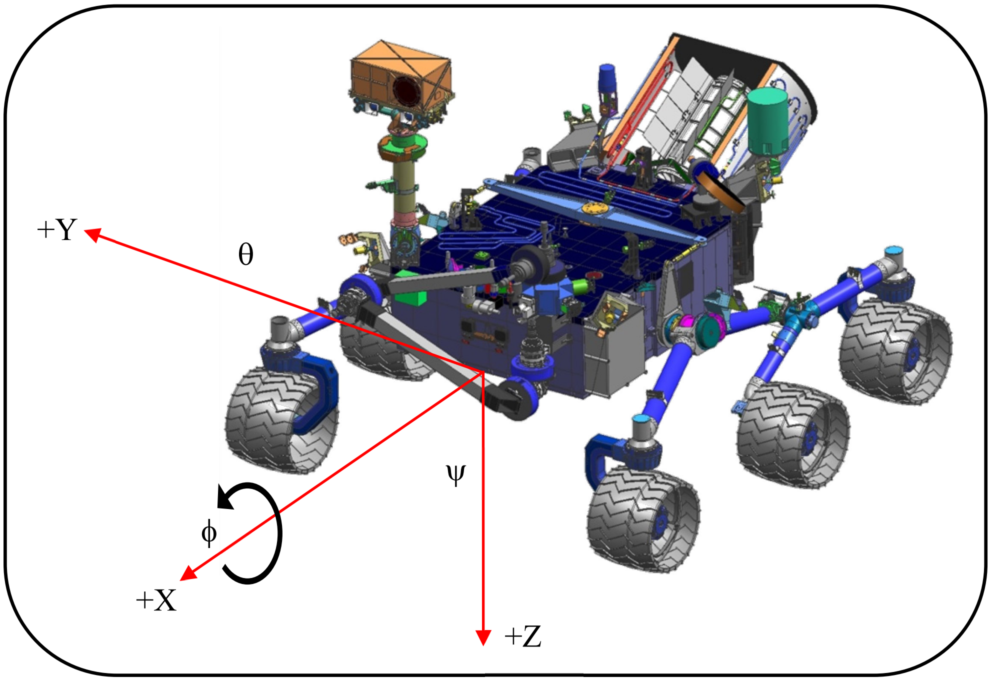
Figure 5 Rover Navigation (RNAV) Coordinate Frame



The Rover Nav frame is specified via an offset from the current Site frame, and a quaternion that represents the rotation between the two. A new instance of the Rover Nav frame, with a potentially unique offset/quaternion, is created every time the ROVER_MOTION_COUNTER increments.

Orientation of the rover (and thus Rover Nav) with respect to Local Level or Site is also sometimes described by Euler angles as shown in Figure 6, where ψ is heading, θ is attitude or pitch, and ϕ is bank or roll.

Figure 6 Yaw, Pitch and Roll Definitions



5.3.2 Rover Mechanical (Rover Mech) Frame

The Rover Mechanical (RMECH) frame is oriented identically to the Rover Nav frame. The origin is forward of Rover Nav by x=0.09002 meters. In other words, given a point expressed in Rover Mech, if you add (0.09002, 0.0, 0.0) you will get the same point expressed in Rover Nav. Rover Mech is not used by any nominal products (EDR or RDR) but could appear in certain special products, generally having to do with arm kinematics.

5.3.3 Local Level Frame

The Local Level frame is coincident with the Rover Nav frame, i.e. they share the same origin at all times. The orientation is different, however. The +X axis points North, +Z points down to nadir along the local gravity vector, and +Y completes the right-handed system. Thus the orientation matches the orientation of Site frames.

Local Level frames are defined by an offset from the current Site frame, with an identity quaternion.

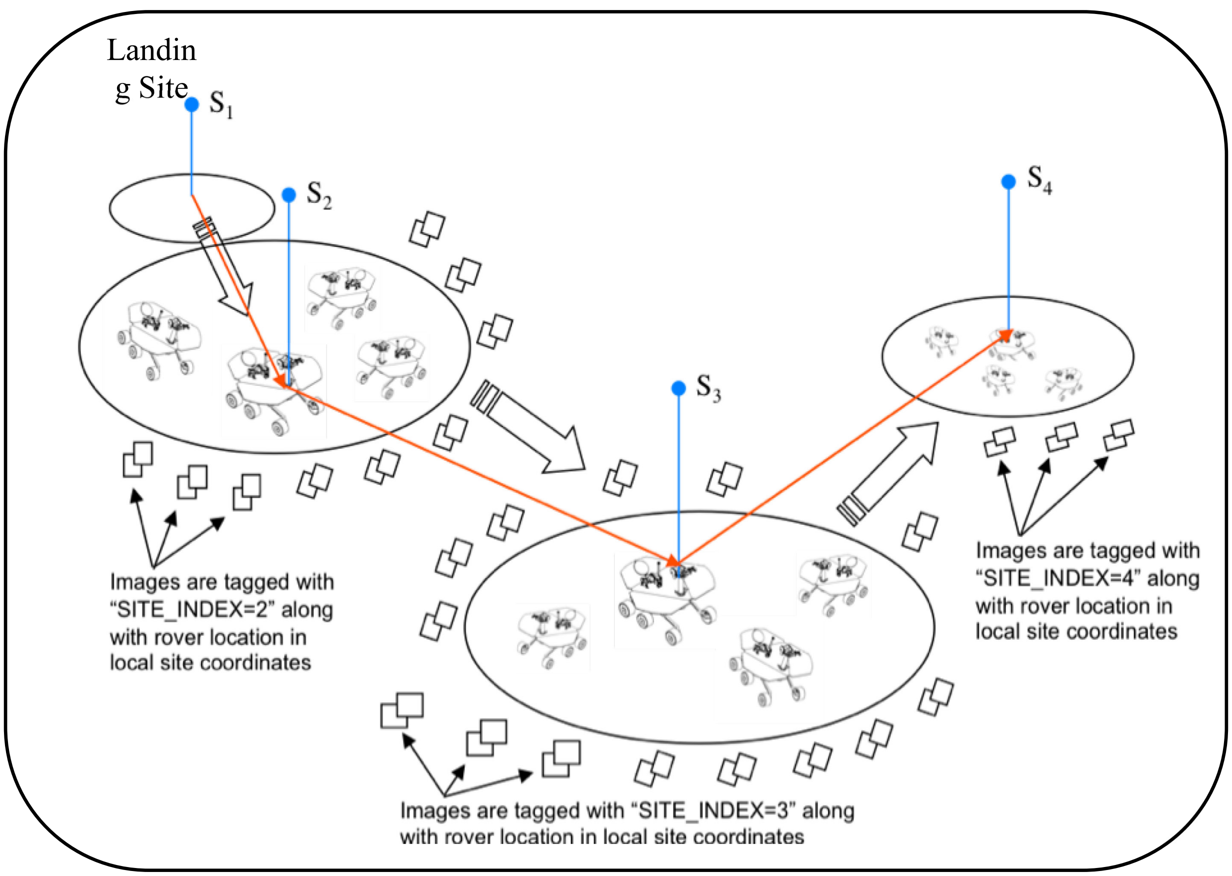
5.3.4 Site Frame

Site frames are used to reduce accumulation of rover localization error. They are used to provide a common reference point for all operations within a local area. Rover Nav and Local Level frames are specified using an offset from this origin. When a new Site is declared, that becomes the new reference, and the offset is zeroed. In this way, long-term localization error is relegated to the offset between Sites, becoming irrelevant to local operations, because the positions are reset with each new Site.

When a Site frame is declared, it is identical to the Local Level frame, sharing both orientation and position. However, the Site frame is fixed to the Mars surface; when the rover moves, Local Level moves with it but Site stays put. Therefore, for the Site frame, +X points North, +Z points down to nadir along the local gravity vector, and +Y completes the right-handed system.

Sites are indexed, meaning there are multiple instances. Site 1 by definition represents the landing location. New Sites are declared as needed during operations, as the rover moves away from the local area. See Figure 7.

Figure 7 Site and Rover Frames



5.3.5 Arm Frames

The frame representing the currently selected arm tool is reported in the arm coordinate system group. The selected tool, given by ARTICULATION_DEV_INSTRUMENT_ID, is arbitrary for any given image and may be surprising; for example SHERLOC-WATSON may not be the selected tool for a SHERLOC-WATSON image. The various tool frames are attached to and aligned with the tool in some manner specific to that tool. See the PPPCS, Ref [7], for actual frame definitions.

6. Applicable Software

The SHERLOC Instrument team designed and developed software tools, or product generation executables (PGEs), to extract, interpret, and output the internal calibration data, spectral position, and spectral intensity L1 and L2 data products. L3-L5 data products will be generated from organic/mineral library comparisons, spatial maps, and organic/mineral region analysis. The SHERLOC team also provides the L0 EDR parsing PGE to Mission System IDS for interpreting the spectroscopy raw binary telemetry to human-readable ASCII text, with additional interpretation on the EDR data products.

6.1 Utility Programs

The table below lists the core software tools that are used to process SHERLOC RDRs through the instrument team data processing pipeline, with the exception of Spectrum Unpack.

Table 8 Key Software Tools for SHERLOC EDR/RDR Processing

Data Type	Software Name	Description	Heritage	Owner
Spectroscopy EDRs	Spectrum Unpack	Parse raw telemetry binary file, separate “active” and “dark” spectra	New	SHERLOC (JPL)
RDRs	Process_Data Unpack	Decode Process_Data EDRs to determine additional necessary processing to generate RDR products	New	SHERLOC (JPL)
	Gain Correction	Apply gain correction to spectra region for consistent scaling	New	SHERLOC (JPL)
	Cosmic Ray Correction	Identify cosmic rays, replace cosmic ray spikes with average or interpolation between points	New	SHERLOC (JPL)
	Dark Subtraction	Subtract dark spectra from active spectra	New	SHERLOC (JPL)
	Wavelength Correction	Convert the spectrum from the pixel domain to the wavelength domain	New	SHERLOC (JPL)
	Normalize Spectrum	Normalize the spectra to the laser photodiode variability	New	SHERLOC (JPL)
	Calibration Comparison	Compare calibration spectra to the primary AIGaN peak	New	SHERLOC (JPL)

	XY Conversion	Convert the scanner azimuth and elevation angles to positions in the target frame	New	SHERLOC (JPL)
	Map Conversion	Calculate the integrated intensity within each specified wavelength region, convert intensity to pixel intensity, and interpolate between xy position points to generate maps	New	SHERLOC (JPL)
	Image Correction	Correct ACI/WATSON images, if necessary (unused in the initial release)	New	SHERLOC (JPL)
	Image Comparison	Determine the homography transformation between each of the ACI images and output the transformation data (e.g. pixel offset)	New	SHERLOC (JPL)
	Image Map Overlays	Output spectrum intensity image laser positions with localized integrated intensity data	New	SHERLOC (JPL)
Autolook	Autolook Generator	Generate Autolook products	New	SHERLOC (JPL)
Engineering Reports	Engineering Unpack	Parse raw telemetry engineering binary file to extract instrument H&S data	New	SHERLOC (JPL)

6.2 Applicable PDS Software Tools

SHERLOC imaging data products, from the ACI and WATSON subsystems, can be viewed using Marsviewer, a display tool developed by JPL/MIPL. It is available in Java, Web and iOS versions from the PDS Imaging Node web site at: https://pds-imaging.jpl.nasa.gov/

They can also be viewed in any other PDS4 compatible imager viewer. The PDS4 viewer (https://sbnwiki.astro.umd.edu/wiki/PDS4_Viewer) offers another option to display PDS4-labeled images and tables. SHERLOC Spectroscopy RDR in ASCII tables may be read using text editors and spreadsheet programs.

Appendix A – SHERLOC Science Data Products

The Instrument Science Data System (SDS) is a ground data system comprised of tools and services to manage and process Mars 2020 instrument data generated during cruise and surface operations phases. Raw data products generated by the rover are transmitted to the ground via the Deep Space Network, where they are reassembled into binary science data frames. Using a series of Product Generated Executables (PGEs), the instrument SDS generates raw science and engineering data products (Engineering Data Records: EDRs) as well as processed data products (Reduced Data Records: RDRs) for analysis by the science team. SHERLOC data are autonomously process and package data products for archive in the Geosciences Node of the Planetary Data System for access by the broader planetary science community. Table 9 provides a list of all RDRs that will be produced by the SDS.

Table 9 A List of RDR Products Archived in the PDS

Product	Description
Processed Spectra	Science and calibration target spectral data that has been processed to remove cosmic rays, subtract dark frames, correct for detector gain variability, and normalize to laser intensity. A table of wavelength values associated with each CCD pixel is also included in this product.
Drift-Corrected Laser Shot Locations	Laser shot positions in the target frame and the image frame, corrected for robotic arm drift
Spectrum Intensity Maps	Up to 36 configurable spectral intensity tables representing Raman or fluorescence spectral variability.
Laser Intensity Map	A table representing the photodiode laser intensity variability within a measurement.
Corrected WATSON Image	WATSON image geometrically and radiometrically corrected.
Corrected ACI Image	ACI image geometrically and radiometrically corrected.

Appendix B – SHERLOC Science RDR Data Structure

Product ID	Description	Tables Objects
RRA	Non-background-subtracted spectral data from the active frame of a target (non-internal calibration). Generated following cosmic_ray correction, or, if the process_data on-board algorithm was applied and background subtraction was not commanded.	COSMIC RAY-CORRECTED SPECTRA REGION 1 COSMIC RAY-CORRECTED SPECTRA REGION 2 COSMIC RAY-CORRECTED SPECTRA REGION 3 --or-- PROCESS DATA SPECTRUM REGION 1 PROCESS DATA SPECTRUM REGION 2 PROCESS DATA SPECTRUM REGION 3
RRB	Cosmic ray-corrected spectral data from the dark frame of a target (non-internal calibration)	COSMIC RAY-CORRECTED SPECTRA REGION 1 COSMIC RAY-CORRECTED SPECTRA REGION 2 COSMIC RAY-CORRECTED SPECTRA REGION 3
RCA	Cosmic ray-corrected spectral data from the active frame of the internal calibration target	COSMIC RAY-CORRECTED SPECTRA REGION 1 COSMIC RAY-CORRECTED SPECTRA REGION 2 COSMIC RAY-CORRECTED SPECTRA REGION 3
RCB	Cosmic ray-corrected spectral data from the dark frame of the internal calibration target	COSMIC RAY-CORRECTED SPECTRA REGION 1 COSMIC RAY-CORRECTED SPECTRA REGION 2 COSMIC RAY-CORRECTED SPECTRA REGION 3
RAR	Cosmic ray list, specifying the spectrum number, channel number, and original intensity value of cosmic ray candidates for the active frame of a target (non-internal calibration)	COSMIC RAY LIST REGION 1 COSMIC RAY LIST REGION 2 COSMIC RAY LIST REGION 3
RBR	Cosmic ray list, specifying the spectrum number, channel number, and original intensity value of cosmic ray candidates for the dark frame of a target (non-internal calibration)	COSMIC RAY LIST REGION 1 COSMIC RAY LIST REGION 2 COSMIC RAY LIST REGION 3
RAC	Cosmic ray list, specifying the spectrum number, channel number, and original intensity value of cosmic ray candidates for the active frame of an internal calibration target	COSMIC RAY LIST REGION 1 COSMIC RAY LIST REGION 2 COSMIC RAY LIST REGION 3

RBC	Cosmic ray list, specifying the spectrum number, channel number, and original intensity value of cosmic ray candidates for the dark frame of an internal calibration target	COSMIC RAY LIST REGION 1 COSMIC RAY LIST REGION 2 COSMIC RAY LIST REGION 3
RRS	Dark-subtracted spectral data of a target (non-internal calibration), with processing indicated by the instrument-specific characters (wavelength calibration, gain correction, laser normalization)	DARK-SUBTRACTED SPECTRA REGION 1 DARK-SUBTRACTED SPECTRA REGION 2 DARK-SUBTRACTED SPECTRA REGION 3 --or-- GAIN-CORRECTED SPECTRA REGION 1 GAIN-CORRECTED SPECTRA REGION 2 GAIN-CORRECTED SPECTRA REGION 3 --or-- WAVELENGTH (NM): WAVELENGTH-CORRECTED SPECTRA REGION 1 WAVELENGTH-CORRECTED SPECTRA REGION 2 WAVELENGTH-CORRECTED SPECTRA REGION 3 --or-- WAVELENGTH (NM) LASER-NORMALIZED SPECTRA REGION 1 LASER-NORMALIZED SPECTRA REGION 2 LASER-NORMALIZED SPECTRA REGION 3 --or-- PROCESS DATA SPECTRUM REGION 1 PROCESS DATA SPECTRUM REGION 2 PROCESS DATA SPECTRUM REGION 3
RCS	Dark-subtracted spectral data of the internal calibration target, with processing indicated by the instrument-specific characters (wavelength calibration, gain correction, laser normalization)	DARK-SUBTRACTED SPECTRA REGION 1 DARK-SUBTRACTED SPECTRA REGION 2 DARK-SUBTRACTED SPECTRA REGION 3 --or-- GAIN-CORRECTED SPECTRA REGION 1 GAIN-CORRECTED SPECTRA REGION 2 GAIN-CORRECTED SPECTRA REGION 3 --or-- WAVELENGTH (NM): WAVELENGTH-CORRECTED SPECTRA REGION 1

		WAVELENGTH-CORRECTED SPECTRA REGION 2 WAVELENGTH-CORRECTED SPECTRA REGION 3 --or-- WAVELENGTH (NM): LASER-NORMALIZED SPECTRA REGION 1 LASER-NORMALIZED SPECTRA REGION 2 LASER-NORMALIZED SPECTRA REGION 3
RCC	Peak fit results to prominent peaks from calibration spectral data, trended over time	CALIBRATION_FIT
RLC	PNG file, converted from input image. <not archived>	<image file>
RPM	Image filename traceability <not archived>	IMAGE_FILENAME_TRACEABILITY
RLD	Image transformation data resulting from robotic arm thermal drift correction <not archived>	IMAGE_TRANSFORMATION
RLP	The position of each laser shot in mm in the target space <not archived>	LASER_SHOT_POSITIONS COMMANDED_LASER_SHOT_POSITIONS
RMx	The integrated intensity for each laser shot using the spectral region defined by the PGE configuration file.	RMx_INTEGRATED_INTENSITY
RLI	The average laser photodiode intensity at each position	LASER PHOTODIODE INTENSITY MAP
RLS	The line (i or x pixel) and sample (j or y pixel) number for each laser shot position on each ACI image, corrected to account for robotic arm drift.	LASER SHOT POSITION
RMO	A summary of image name, laser shot positions, and integrated intensity for each wavelength range at each point.	LASER_SHOT_POSITIONS WAVELENGTH_REGIONS SPECTRAL_INTENSITY