esac

European Space Astronomy Centre
P.O. Box 78
28691 Villanueva de la Cañada
Madrid
Spain
T +34 91 8131 100
F +34 91 8131 139
www.esa.int

DOCUMENT

Metadata Definition for Solar Orbiter Science Data

Prepared by	**Helen Rose Middleton**
Reference	**SOL-SGS-TN-0009**
Issue	**2**
Revision	**6**
Date of Issue	**16 Apr 2024**
Status	**Ready for Approval**
Document Type	**TN**
Distribution	**PIs, CoIs, SOC, Project Scientists**

APPROVAL

Title Metadata Definition for Solar Orbiter Science Data	
Issue 2	**Revision** 6
Author [Anik De Groof, Andrew Walsh, David Williams, in collaboration with Solar Orbiter MADAWG] Helen Middleton, Archive Scientist	**Date** 16 Apr 2024
Approved by	**Date**
Project Scientist, Daniel Müller	Daniel Mueller Digitally signed by Daniel Mueller Date: 2024.04.16 16:32:21 +02'00'
Mission Manager, Anik De Groof	Anik De Groof Digitally signed by Anik De Groof Date: 2024.04.16 16:14:10 +02'00'
SWT	Approved in SWT#38, see minutes.

CHANGE LOG

Reason for change	Issue	Revision	Date
First Issue released at SOWG#6 for review by instrument teams	1	0	13/01/2015
Second Complete Issue after Instrument team feedback	2	0	29/06/2015
Update after comments during SOWG#7	2	1	21/07/2015
Minor update containing clarification to CDF standard	2	2	23/07/2015
Minor updates and typo fixes	2	3	24/09/2018
Update of recommended CDF version to 3.7. Small clarifications & error fixes	2	4	02/09/2019
Clarifications throughout, some keywords now mandatory or optional	2	5	01/12/2022
Minor changes plus light travel time correction text.	2	6	26/06/2023

CHANGE RECORD

Issue 2	**Revision** 6		
Reason for change	**Date**	**Pages**	**Paragraph(s)**
Approvers changed from PIs to PS, MM and SWT.	19/03/2024	2	
Any language inferring that the requirements within this document are proposals has been removed.	26/06/2023	All	All
Added note that JP2 files should contain the same metadata as the L2 files they depict but with an accurate filename keyword value.	26/06/2023	11	2.1.2
Descriptor restrictions updated to be more specific: lower-case and only normal characters, i.e., no special characters. Note: the hash character (#) was briefly allowed but now withdrawn.	26/06/2023	13	2.1.3 (Table 2-2)
Added the sentence to the text describing the use of the BLANK keyword: "If there are no undefined pixels, then the BLANK keyword is not required."	18/03/2024	17	3.1.1

Added David Berghmans' changes to the OBJECT (p24) and TARGET (p29) accepted keyword values.	16/04/2024	24 & 29	3.1.1.3 (Table 3-3) 3.1.1.4 (Table 3-4)
If there are no undefined pixels in the image, then the BLANK keyword is not required to be present.	17/03/2024	30-31	3.1.1.5 (Table 3.5)
Alternative text inserted (written by Bill Thompson, Jesper Schou and Bogdan Nicula) to replace the incorrect paragraph: 'WCS keywords should not include any corrections for light travel time and stellar aberration…'	18/03/2024	33	3.1.1.8
Paragraphs (written by Bill Thompson, Jesper Schou and Bogdan Nicula) inserted into Section 3.1.1.9 to explain how the CRLN_OBS keyword value should be corrected for light travel time.	18/03/2024	38	3.1.1.9
Description for CRLN_OBS adapted to request light travel time correction: 'corrected for light travel time (see above).'	18/03/2024	40	3.1.1.9 (Table 3.9)
Note added that the EXTEND keyword is not expected in an IMAGE extension.	26/06/2023	45	3.1.2.1.1
Instrument attribute is now the instrument and the sensor, since otherwise there is no place for the sensor to be specified	06/10/2023	56	3.2.2.1 (Table 3-16)
Instructions to abbreviate PI_name and PI_affliation removed.	17/03/2024	56	3.2.2.1 (Table 3-16)
CDF global attribute HTTP_LINK made optional, in line with FITS keyword INFO_URL	26/06/2023	57	3.2.2.1 (Table 3-16)
Format specified for TIME_MIN and TIME_MAX as being ISO	11/10/2023	59	3.2.2.3 (Table 3-18)
For variable attribute UNITS: use the symbols not full words, e.g., T not Tesla, or m/s not metres/second. For variables with no units (e.g., a ratio or a direction cosine) use 'unitless'.	04/10/2023	62	3.2.3.1 (Table 3-20)
For variable attribute VAR_NOTES: can be 'none' if no extra details can be added, e.g., for EPOCH.	05/10/2023	63	3.2.3.1 (Table 3-20)
For variable attribute SI_CONVERSION: For variables already given in SI units 1.0>[SI unit], e.g., 1.0>s. For unitless variables, 1>unitless	04-10-2023	64	3.2.3.2 (Table 3-21)

Table of contents:

1 INTRODUCTION

Most Solar Orbiter science goals are based on coordinated observations of its 6 remote-sensing and 4 in-situ instruments [SolO-SRD]. As a direct consequence, the scientific users of Solar Orbiter data will need to access data sets from different instruments and perform scientific analysis on a multi-instrument dataset. Due to the mission's particular orbit and operational constraints, also data products from other space or ground assets might be required to provide the necessary context for the Solar Orbiter data.

However, the scientist/data user in question might not be an expert on all the different types of instruments and data. Connection science in particular requires the combination of very different types of data and the interoperability of analysis software of two quite distinct scientific communities.

It is therefore crucial for the scientific outcome of the mission, and in the benefit of the whole scientific community, to make the full range of Solar Orbiter data easily accessible and easily interpretable by defining data products which are consistently formatted in a transparent and standard manner. All data (both operational and science) will also be stored in the mission archive (SOAR) at the Solar Orbiter SOC (ESAC, Madrid) from which (a subset) will be made available to the science community.

Another important aspect for data provision to the scientific community is the compliance to existing tools and broadly used software. As an answer to the complementarity of recent missions like STEREO, SDO, IRIS, etc. the solar physics community has developed over the last years a wide range of common data archives, retrieval systems and analysis software tools that facilitate the combination of many data types and sources. Therefore, it is highly recommended to also make the Solar Orbiter science data compatible with the most widely used software packets, so that data can be digested with tools like (J)Helioviewer, 3Dview, AMDA, 3Dview, FESTIVAL and others.

To facilitate connection of in-situ and remote-sensing observations, compliance with connection models, e.g., magnetic field extrapolations like PFSS, radial constant speed outflow models, and (CME) propagation models like ENLIL, is highly recommended.

In this technical note, we describe the required format for Solar Orbiter's data products and metadata.

1.1 Reference Documents

[ANCDATA]	SOL-SGS-TN-0017 SOC-Provided Ancillary Data For Solar Orbiter
[CDF]	CDF User's Guide (V3.7.1), http://cdaweb.gsfc.nasa.gov/pub/software/cdf/doc/cdf371/cdf371ug.pdf
[DPAICD]	SOL-SGS-ICD-0002, Solar Orbiter Data Producer to Archive ICD
[EID-A]	Experiment Interface Document – Part A, Iss. 5.0, SOL.EST.RCD.0050_05_00
[FITSpaper]	Definition of the Flexible Image Transport System, Iss. 4.0, https://fits.gsfc.nasa.gov/standard40/fits_standard40aa-le.pdf
[FITS+WCS]	FITS World Coordinate System (WCS), http://fits.gsfc.nasa.gov/fits_wcs.html
[FITSchecksum]	Checksum Keyword Convention, http://fits.gsfc.nasa.gov/registry/checksum/checksum.pdf
[IAU15]	IAU resolutions, 2015 general assembly: https://www.iau.org/static/resolutions/IAU2015_English.pdf
[IOR ICD]	SOL-SGS-ICD-0003, Solar Orbiter Instrument Operations Request ICD, Iss 1.3
[ISTP]	ISTP guide, http://spdf.gsfc.nasa.gov/istp_guide/istp_guide.html
[LLdata]	SOL-SGS-TN-003, Solar Orbiter Low-Latency Data: Concept and Implementation
[LLCDFICD]	SOL-SGS-ICD-0004, Solar Orbiter ICD for LL Data CDF Files
[LLFITSICD]	SOL-SGS-ICD-0005, Solar Orbiter ICD for LL Data FITS Files
[SAP]	SOL-EST-PL-8539, Solar Orbiter Science Activity Plan, Iss. 0.1
[SIRD]	SOL-EST-RS-4514, Solar Orbiter Science Implementation Requirements Document
[SolO-SRD]	SOL-EST-RS-1858, Solar Orbiter Science Requirements Document
[WCSpaper]	Coordinate systems for solar image data, W. Thompson, http://adsabs.harvard.edu/abs/2006A%26A...449..791T
[WCSpaper2]	Precision effects for solar image coordinates within the FITS world coordinate system, W. Thompson, http://adsabs.harvard.edu/abs/2010A%26A...515A..59T (defines some additional keywords)
[WCStime]	WCS paper on time coordinates, A. Rots, http://hea-www.cfa.harvard.edu/~arots/TimeWCS/WCSPaper-IV-v1.1A4.pdf

More papers/webpages that can be of interest:

[SECCHI FITS]	http://hesperia.gsfc.nasa.gov/ssw/stereo/secchi/doc/FITS_keywords.pdf
[SWAP FITS]	http://proba2.sidc.be/data/SWAP/level0
[EIS FITS]	http://solarb.uio.no/software/eis_userguide/fits_kw_list.htm
[IRIS FITS]	https://www.lmsal.com/iris_science/doc?cmd=dcur&proj_num=IS0077&file_type=pdf

1.2 List of Tables

2 SOLAR ORBITER DATA PRODUCTS AND FORMATS

As explained in the introduction, it is of crucial importance for the science goals of Solar Orbiter to facilitate scientific analysis of multi-instrument and multi-mission datasets. This requires a minimal set of science data formats that are common to all Solar Orbiter data products and are consistent with commonly used data formats in the science community.

2.1 Data Formats and File Names

2.1.1 Solar Orbiter Data Formats

The formats for the Solar Orbiter scientific data products are as follows:

- Image data should be stored in FITS files with standardized FITS keywords (see [FITSpaper] and Sect. 3.1). One FITS file can contain one or a series of images.
- Time series from in situ instruments should be stored in CDF files, one per day, or one per burst mode period; CDF files should be version 3.7 or greater. 3.7 is recommended for Solar Orbiter by the Space Physics Data Facility. Metadata should follow the standard described in Sect. 3.2 CDF Metadata Standard. Time series from remote sensing instruments can be stored in FITS files, within an appropriate binary table extension (e.g., STIX and Metis lightcurves).
- Combined data sets or spectral data (e.g., STIX, SPICE) can be stored in FITS files with binary table extensions or in multi-image FITS files. The primary FITS header should be standardized according to the rules set out in Sect. 3.1. FITS Metadata Standard.
- All science data should include a header; FITS keywords; or CDF global attributes, variable attributes and support variables that follow the Solar Orbiter standard described in this document. The use of detached headers (metadata in a separate file) is not acceptable.

Higher processing levels of data products might need different formats and are defined below.

2.1.2 Definition of Processing Levels

Historically, data processing levels have been defined differently in different scientific communities. However, it would be highly beneficial to use a consistent definition throughout the Solar Orbiter data products.

The definition below is based on the high-level description of [EID-A] and [SIRD] but adds more detail in the description of the data levels. Due to the specific nature of some of the SolO data products, Instrument Teams (IT) are free to define multiple types of data with each data level, if appropriate. These should be distinguishable through the filename (see Section 2.1.3). All data levels available for each instrument should be stored in the mission archive (SOAR) at SOC, although not all data processing levels will necessarily be open to the scientific community. Optional data products are formatted *italic* in the table below.

Note that the L0-3 level data are described as ‘science’ data.

Table 2-1 Solar Orbiter Data Processing Levels

Level	Source	Data Type	Format and Metadata content
L0	IT	"Raw" data, unpacked and decompressed data	Data format preferable in FITS or CDF but can be different if this is more appropriate. Metadata reflect the information that was available in the TM packets only.
L1	IT	"Engineering" data, uncalibrated	FITS or CDF, metadata follows Solar Orbiter standard for L1 (see Section 3): *Note that this level might not always apply to instruments using a complete processing and calibration pipeline onboard, like PHI.*
L2	IT	"Calibrated" data, science quality	FITS or CDF, metadata follows Solar Orbiter standard for L2 (see Section 3): full attitude information in WCS coordinate frame and time in UTC.
L3	*IT*	*Higher-level data*	*Data format as appropriate. The format of Level-3 data, calibration data and ancillary data can be chosen depending on the type of data product and the objectives. However, as much as possible standard formats should be used (MPEG, FITS, JPEG2000, CDF, PNG, ...).*
CAL	IT	Calibration data	Data format as appropriate. *Not all calibration data are necessarily open to the scientific community.*
ANC	IT/SOC	Ancillary data	Data format as appropriate. *Not all ancillary data are necessarily open to the scientific community.*
LL01	SOC	LL engineering data, output of LL pipeline	FITS or CDF, metadata follows Solar Orbiter standard, with some specifics for LL01 data (see [LLFITSICD] and [LLCDFICD] : time in OBT, attitude in instrument detector reference frame.
LL02	SOC	Operational LL data, enhanced with S/C HK	FITS or CDF, metadata follows Solar Orbiter standard, with some specifics for LL02 data (see [LLFITSICD] and [LLCDFICD]: time in UTC, attitude in WCS coordinate frame.
LL03	SOC	Visualisation of operational LL data, in "quicklook" format	'Quicklook' data in PNG or JPEG2000 (details TBC). This level is also used for LL data products derived from (multiple) LL02 products.

Note that we do not specify a level for LL TM that has been fully processed and calibrated by the instrument team. These should be classified as 'L2'. Higher level, derived data products are part of 'L3' data.

The data levels can be described in more detail as follows:

- Level 0 data (L0) are basically telemetry packets, that are decommutated and decompressed and formatted in standard formats which may be different from CDF or FITS if more appropriate[1]. The metadata should contain only the information that was available from the packet headers.

[1] FITS format is preferred for all imagery though (recommendation RSWG on 6 July 2015).

No metadata are added from elsewhere. L0 data may be structured per sensor or instrument unit, similar to how the data were packed onboard.

- Level 1 data (L1) are still uncalibrated data. These data are formatted in standard formats, i.e., FITS files for images (1 FITS file per image, image sequence or spectral cube) and CDF (daily files or per burst period) for time series. They form the basis of the scientific quality data in Level 2. They are still expressed in engineering units (counts) and may contain extra engineering metadata from other sources. Still, this type of data is not yet of science quality. It may also lack details on the instrument's boresight or attitude at the time of the data acquisition. Examples are "raw" images in counts, not corrected for flatfield, dark current, etc.; particle spectra in counts; magnetic field data in which S/C disturbances are still included; etc.

  In case Level 1 data are distributed to the scientific community, they should be accompanied by calibration files and processing software to enhance them to Level 2 for scientific analysis.

- Level 2 data (L2) are calibrated data, ready for scientific analysis. Depending on the type of instrument, "full calibration" might have a different meaning. Data products for which this is appropriate, instrument counts/measurements will be calibrated to physical units. For other instruments (e.g., broadband imagers) it might be more appropriate to limit calibration to the correction of data for all known instrumental defects, and calibrate detector counts in exposure-normalized counts rather than in physical units. In any case, level 2 data should be the reference science product for the scientific user, shielding the user from any instrument peculiarities. All the available orbital and attitude information is applied, and coordinates expressed in scientific coordinate systems (WCS for FITS).

  Level 2 data are likely to contain subcategories, depending on the science goal. Time series for example may be distributed both in full temporal resolution as resampled to lower cadence. These data should however be suitable for quantitative analysis, otherwise they should be categorized as Level 3 data.

- Level 3 data (L3) are higher-level data products, in the sense of derived data products that either combine several 'single' Level-2 data products or are a derived data product that is the result of data analysis. Examples of the former are: movies for imagers; images that combine FOVs of different sub-units; mosaics; graphical plots of dynamic spectra; J-maps; parameters calculated from multiple instruments' data, e.g. plasma beta, Alfvén Mach number. Images of Level-2 products, e.g., JP2 files, that can include the same metadata as the Level-2 files, but the filename keyword value should still be accurate. Examples of derived, value-added data products are: event catalogues; graphs/plots linking in-situ to remote-sensing measurements. Many of the data products in this category may be only defined after launch. Level 3 data are not necessarily limited to one instrument.

- Calibration data (CAL) are data used typically to derive the Level 2 data from the Level 1 data. The calibration data could be generated from any available data source, incl. pre-launch test data, spacecraft attitude information, some Level 0/1 data (e.g., dark images) and appropriate spectrum synthesis models. The format of the calibration data is chosen depending on the type and objective of the data product but should re-use as much as possible the data formats used for Level 1-3 data.

- Ancillary data (ANC) are any data that are provided to the scientific community as context for the scientific data. These are data products derived from S/C or payload HK TM that are useful for scientific analysis, and could include: instrument status for all payload, SPICE kernels for orbit and S/C attitude, time conversion files (OBT to UTC), time periods of burst modes, science operation plans, etc. The data format will depend on the type of ancillary data. MADAWG recommends to also include instrument-specific ancillary data, in the form of an 'engineering logbook' that accompanies each dataset and lists the available engineering information such as special observing modes, anomalies, data gaps, etc.

- Low-Latency data (LL01 or LL02) are a small subset of scientific data that are downlinked with highest priority. In general, they are either a subset of the nominal science data, or a special data product with typically lower time and/or spatial resolution and lower quality than the scientific datasets. LL data are operational in nature in the sense that they are necessary for mission planning and S/C pointing definitions at the SOC. These data products are time critical and are therefore produced at the SOC: LL TM is first reformatted by instrument-provided LL data pipelines into Level LL01 data, and then post-processed by SOC software to apply common operations (like conversion to UTC, application of the spacecraft attitude, and known instrument alignments), resulting in Level LL02 data. SOC will also produce a quick look format of the latter, LL03 data, for visualisation in the Low-Latency Data Visualisation tool. All 3 SOC-produced 'operational LL data' will be available through the mission archive, at least to the instrument teams. Because of the timely operational nature of these products, subsequent reprocessing is excluded, except in case of recovered data gaps.

Note that during times of low downlink performance, LL data might be the only recent data available from Solar Orbiter during several months. Therefore, instrument teams may choose to produce an enhanced, better calibrated, version of the LL data starting from the same LL01 TM packets. These will however not be labelled as Level LL but rather as Level-2 or Level-3 data.

This document gives the metadata required for the science files; any requirements for Low Latency metadata are described in the relevant ICD ([LLFITSICD] or [LLCDFICD]). The levels at which metadata is mandatory/optional may be specific (e.g., L1) or more inclusive (e.g., L2+, which would apply to L2 and L3).

2.1.3 Solar Orbiter File Naming Convention

The naming convention described below shall apply to all Solar Orbiter data product files, including calibration and ancillary data. It is based on both the ISTP file naming convention used in the in-situ community and the needs and common practices in the remote-sensing community. As a compromise between the needs and conventions of the in situ and remote sensing payloads, Solar Orbiter data files will be named according to the following convention, which has been defined with the aim of providing different instrument teams with the flexibility they need within a consistent framework:

`source_level_descriptor_datetime_version_freefield.extension`

Where the different fields that make up the filename are separated by underscores and are defined as follows:

Table 2-2 Fields in the Solar Orbiter file naming Convention

Field	Description	Mandatory / Optional
source	Identifies the data as from Solar Orbiter. Content will be 'solo'.	Mandatory
level	Processing level. Content will either be L0, L1, L2, L3, LL01, LL02, LL03, ... according to the Solar Orbiter data processing level standard in Sect. 2.1.2.	Mandatory
descriptor	The instrument (mandatory), and if appropriate the sensor/detector and/or data product, each separated by a hyphen. Including the instrument is mandatory. Detector and data product information can be included (or not) as required, e.g., metis, swa-eas-pad or epd-flux are all valid descriptors. However, the descriptor must **uniquely identify the type of data in the file**, for example EUI FSI images are taken in different wavelengths and so the descriptor needs to specify the wavelength (i.e., eui-fsi cannot be a descriptor on its own, it must be e.g., eui-fsi304). The descriptor should only use lower-case letters or numbers and use no special characters. Level 3 data derived from multiple instruments will have 'multi' as the first part of descriptor. The meaning of the descriptor for ancillary (and planning) data is slightly different, i.e., <creating actor> + '-' + <product>	Mandatory
datetime	The timestamp of the data in the file, formatted according to Section 2.1.3.1 below, corresponding to FITS keywords OBT_BEG (L0) or DATE-BEG (L1+) or CDF keyword TIME_MIN as the time at which the observation was made by the instrument/spacecraft. If OBT_END or DATE-END (CDF: TIME_MAX) are applicable, these should be appended to the OBT_BEG or DATE-BEG, with a hyphen, to this datetime field.	Mandatory
version	The version of the file, formatted as capital V + 2-character incremental number padded with 0, e.g., V01 Note that this is different for LL files: see [LLCDFICD] and [LLFITSICD] for details.	Mandatory
free field	An optional field of which the format is free, but decimal points or underscores cannot be used. **Free fields are intended for internal instrument team use only and will be ignored upon ingestion into the Solar Orbiter Archive at ESAC** – the filename will not change.	Optional
extension	File extension: '.fits' for all FITS files, '.cdf' for all CDF, '.jp2' for JPEG2000 files (more extensions might be added later). Simple ASCII text files must also have an extension '.txt'.	Mandatory

2.1.3.1 Datetime Format

1. The datetime field will be separated from other fields in the filename by an underscore, like the other sections of the filename.
2. Time will be given as OBT for L0 files, and UTC for all higher-level files (L1+).
3. The datetime field will be (at the very least) the start of the observation interval (for time series) or the start of the integration for images (i.e., content of FITS keyword DATE-BEG or OBT_BEG in L0).
4. Granularity of datetime in UTC can be chosen as appropriate, but date will be separated from time with a capital T. Include all fields that apply, e.g.:

   a. For day granularity files: yyyymmdd
   b. For millisecond granularity files: do not use the decimal point but add the required number of digits to the time field: yyyymmddThhmmssxyz

   For L0 data, the datetime field shall contain the coarse OBT of the observation in the file, padded to 10 digits, rather than any UTC timetag. For notes on the format of the OBT coarse and fine parts, see Section 3.1.1.2.

5. An end datetime is optional, but if present shall be separated from the start datetime by a hyphen, '-'.
6. The end datetime shall have the same granularity as start datetime.
7. For time series, in the absence of the end datetime the file duration is assumed to be the granularity of start datetime, e.g., yyyymmdd has 1 day duration, while yyyymmddThh has one hour duration.

In situ normal mode files are encouraged to be daily, without an end datetime. In situ burst mode files are encouraged to be per burst, with an end datetime. For RS imagery, the recommendation is to only specify the acquisition start time unless the observation is constructed from many acquisitions taken over a substantially long period.

2.1.3.2 Version format

The version of a data product links to multiple types of versions linked to e.g., the software producing the product, the version of parameter configuration, the calibration data used to calibrate it, updates in metadata format, etc. The versions of the different 'actors' that touched the data will be described in the metadata (see sections on FITS and CDF Metadata Standard). The filename however will contain a version number that is incremented whenever one or more of the individual versions got updated.
This data product version should be formatted as 'V' + and incremental number padded with '0' to 2 characters. **It is of vital importance that every instrument team keeps track of the link between the version number and the individual version components consistently.**

2.1.3.3 Capitalisation

Filenames should be kept lowercase, except for a few meaningful letters that should always be capitalised: L for level, V for version, T as separator between date and time, and ANC, CAL as 'levels' for ancillary and calibration data respectively.

2.1.3.4 Example Filenames

A Level 2 SoloHI image might have the following filename:

`solo_L2_solohi-1fs_20211127T003100_V01.fits`

A MAG burst **daily** file in RTN coordinates might have the following filename:

`solo_L2_mag-rtn-burst_20220331_V02.cdf`

An EPD L0 file might have the following filename; note that the datetime section is in OBT, padded to 10 digits:

`solo_L0_epd-step-nom_0676080000-0676166399_V01.cdf`

SWA-EAS Pitch Angle Distributions from a burst mode might have the following filename:

`solo_L2_swa-eas-pad_20181012T045630-20181012T050630_V01.cdf`

A highly compressed, yet uncalibrated EUI-FSI 304Å image that got downloaded as Low-Latency data, is complete, and was produced by the SOC might have the following name:

`solo_LL02_eui-fsi304_20201008T121230_V01C.fits`

and can be distinguished from the same uncalibrated image at science resolution that was downloaded much later from the level in the filename. Note that the 'C' is no longer required since the science file must be complete:

`solo_L1_eui-fsi304_20201008T121230_V01.fits`

A SPICE summary map of the FIP fractionation abundance data may be named:

`solo_L3_spice-fip_20201008T121230_V03.fits`

2.1.3.5 Calibration and Ancillary Files

Calibration and Ancillary Data will follow the same standard. Calibration files will have CAL included as the *level*; further subcategories can be added as appropriate in *descriptor*, separated by hyphens. E.g.

`solo_CAL_mag-ibs-offset_20191001-20201001_V01.cdf`

Similarly, ancillary files will have 'ANC' as *level*. SOC-Produced Ancillary files (e.g., the orbit spice kernel) will have 'soc' in the descriptor, e.g.

`solo_ANC_soc-orbit_20190101-20190301_V01.bsp`

Ancillary files are described in detail in [ANCDATA].

3 SOLAR ORBITER METADATA STANDARD

The following sections describe a proposal for a standardized list of metadata. FITS keyword lists are defined in Sect. 3.1 and CDF metadata variables and attributes in Sect. 3.2. CDF Metadata Standard

3.1 FITS Metadata Standard

The following subsections contain a list of keywords that should be provided in Solar Orbiter science FITS files of Level 0 ('raw' data), Level 1 (engineering data) and Level 2 (calibrated science data). The proposal follows the FITS standards [FITSpaper] and use of the World Coordinate System (WCS) [WCSpaper] for position and pointing metadata. The list of FITS keywords is based on the metadata definitions used for the STEREO/SECCHI [SECCHI FITS] and PROBA2/SWAP [SWAP FITS] and also inspired by keyword definitions of Hinode/EIS [EIS FITS] and IRIS [IRIS FITS] instruments.
In the keyword lists below, grouped by metadata type, we distinguish 3 types of keywords:

- M: Mandatory keywords required for compliance to the standard FITS format.
- P: Keywords that are mandatory in the Solar Orbiter FITS files. They should comply with the description and units described below.
- O: Keywords that are optional and can be defined if applicable to the instrument in question. However, if they are defined with the proposed name they should be provided in the units and follow the description below.

Note that keywords marked with * are optional in FITS standard but if they are present in a FITS header, they must be used as defined in the standard.

Any additional instrument specific keywords and values can be included as long as they are referenced in the dedicated dataset [product] description documents (DPDD).

3.1.1 Primary Header Keywords

The keywords defined in the tables below apply in general to all FITS Header Data Units (HDU) that contain a data array. In case a FITS file contains extensions in addition to the primary HDU, additional keywords apply to the extension headers; these are specified in Section 3.1.2.
The FITS header shall comply to the structure and formatting specified in [FITSpaper]. Below we repeat some basic rules:

- FITS header keyword records are 80 characters long and consist of a keyword, a value indicator ('=␣'), a value, and an optional comment (preceded by ' /'). The keywords COMMENT, HISTORY and, if used, CONTINUE, do not need a value indicator or value.
- The keyword name shall be a left justified, 8-character, space-filled, ASCII string with no embedded spaces. Only upper-case letters, numbers 0 to 9 and hyphen '-' and underscore '_' are allowed.

- The structure of the value field depends on the data type of the value: strings should always be quoted, numerical values not.
- In case a keyword is not defined, the keyword will be omitted from the header. This should however never happen in case of mandatory keywords (type ‘M’ or ‘P’ in particular data level).
- The FITS standard forbids NaN as a numerical value in the value of the keywords. If the data is integer (BITPIX is positive), the BLANK keyword should be used to give the value given to pixels with an undefined physical value. If there are no undefined pixels, then the BLANK keyword is not required. If the data is floating point, (BITPIX is negative), NaN is used for undefined pixels, and the BLANK keyword is not used (NaN is not allowed in the header and so the BLANK keyword is omitted).
- Units are specified in the comment field with square brackets at the beginning of the comment, e.g., *WAVELNTH=304 / [Angstrom] wavelength in Angstroms.* See Greisen and Calabretta (2002) for details.

3.1.1.1 Basic Keywords

Table 3-1 Basic FITS Keywords

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
SIMPLE	T	file conforms to basic FITS standards	T (for true)	Boolean	SIMPLE is a logical keyword specifying whether the file conforms to basic FITS standards; Solar Orbiter data should always comply, so T (for True) is the only valid option.	M	All
BITPIX	16	number of bits per data pixel, rounded up	8, 16, 32, 64 for binary integers -32, -64 for single/double precision floating points	Integer	BITPIX states the number of bits per pixel and can only take values 8, 16, 32, 64 for binary integers and -32 or -64 for single, resp. double precision floating points (see table 8 in [FITSpaper]) so the actual bit depth should be rounded up if necessary.	M	All
NAXIS	2	number of axes in data cube	0, ..., 4+, positive integer	Integer	The NAXIS keyword indicates the number of dimensions in the data (0 indicates header only).	M	All
NAXIS1	1024	length of data in axis 1	positive integer	Integer	The NAXISx keyword (where x is the number of dimensions specified in NAXIS), describes the number of pixels along that axis. The example given here (along with NAXIS2) is for a 2D image with 1024x1024 pixel resolution. A 3D data cube would need an extra keyword NAXIS3.	M	All
NAXIS2	1024	length of data in axis 2	positive integer	Integer	Length of the second axis (y, number of rows in 2D image).	M	All
EXTEND	F	F = no extensions T = extension(s)	T, F (for true, false)	Boolean	EXTEND specifies whether the FITS file contains extension tables or multiple images or not, and is proposed to be a standard keyword in SolO data FITS headers. Mandatory keywords in extension table headers are specified in Table 10 of [FITSpaper] and SolO standard keywords in Section 3.1.2.	P	All

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
LONGSTRN	'OGIP 1.0'		'OGIP 1.0'	String	Specify this keyword in case the FITS header contains strings longer than the 68-character limit, to allow continuing the string value over multiple keywords. This continuation convention uses an ampersand character at the end of each substring to indicate that it is continued on the next keyword, and the continuation keywords all have the name CONTINUE without an equal sign in column 9. The string value may be continued in this way over as many additional CONTINUE keywords as is required. See https://fits.gsfc.nasa.gov/registry/continue_keyword.html	O	All

3.1.1.2 General Description Keywords

Table 3-2 General Description FITS Keywords

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
FILENAME	'solo_L2_eui-fsi304-image_20200512T085922556_V04.fits'	FITS filename	See filename convention	String	Name of FITS file. FILENAME should follow the conventions laid out in Sect. 2.1 (Data Formats and File Names), with the date/time consistent with the information in OBT_BEG (L0) or DATE-BEG (L1+) below.	P	All
FILE_RAW	'BatchRequest.PktTmRaw. SOL.0.2018.262.13.26.30.860.jphr @2018.262.13.26.31.239.1.xml'	Raw filename(s)	*TBD*	String	Optional keyword referring to the raw file (reassembled TM packets) the FITS file was derived from. If two TM files are required to build the LL01 FITS file (for example when there is a file for context as well the "target" TM file), then their names should be separated within the single string by a semicolon ";".	O	All

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
PARENT	'solo_L1_eui-fsi304-image_20200512T085922556_V04.fits'	Source file current data product	See filename convention	String	Name of the parent or input file that got processed to the current one. This keyword is proposed to maintain heritage, in combination with FILE_RAW, as in the CDF attribute 'Parents'. The PARENT keyword typically contains the name of the FITS file of 1 level lower than the current.	P	L1+
APID		APID number of associated TM	positive	Integer	APID number of associated TM.	O	All
DATE	'2020-01-27T14:23:57.167'	[UTC] FITS file creation date	any date+'T'+time	String	Date and time of FITS file creation, in UTC, in ISO-8601 format 'yyyy-mm-ddThh:mm:ss.sss'	P	All
DATE-OBS	'2020-01-22T14:00:00.000'	[UTC] Deprecated, same as DATE-BEG.	any date+'T'+time within mission duration	String	Time of observation, in UTC, in ISO-8601 format. Historically, this has usually referred to the start of the observation, though it has sometimes been used for the midpoint. The [WCStime] paper defines two new keywords, DATE-BEG and DATE-AVG, to replace DATE-OBS. To support both older and newer software, Solar Orbiter files will contain both DATE-OBS and the DATE-BEG, which will have the same value.	P	L1+
DATE-BEG	'2020-01-22T14:00:00.000'	[UTC] Start time of observation	any date+'T'+time within mission duration	String	Start time of observation, in UTC, in ISO-8601 format. DATE-BEG replaces DATE-OBS that got depreciated. The value of DATE-BEG is also reflected in the FITS filename.	P	L1+

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
DATE-AVG	'2020-01-22T14:00:30.000'	[UTC] Average time of observation	any date+'T'+time within mission duration	String	Average time of observation, in UTC, in ISO-8601 format. This will usually be the midpoint between the start and end times, although DATE-END is optional. Mandatory for Solar Orbiter as all attitude and WCS metadata are based on this time.	P	L1+
DATE-END	'2020-01-22T14:01:00.000'	[UTC] End time of observation	any date+'T'+time within mission duration	String	End time of observation, in UTC, in ISO-8601 format. If applicable, the value of DATE-END is also reflected in the FITS filename.	O	L1+
TIMESYS	'UTC' or 'OBT'	System used for time keywords	See [FITS+WCS]	String	System used for time keywords. For Solar Orbiter, this should be 'UTC'. See Greisen and Calabretta in [FITS+WCS] for a complete list of possible values.	P	L1+
TIMRDER	0.2	[s] Estimated random error in time values			Estimated random error in time values, expressed in seconds.	O	L1+
TIMSYER	0.1	[s] Estimated systematic error in time values			Estimated systematic error in time values, expressed in seconds. The total error is the root-summed-square of the random and systematic errors.	O	L1+
OBT_BEG	636564608.5277252	start acquisition time in OBT	positive float	Float	Mandatory keyword that contains the start time of data acquisition in onboard time units (double-precision, floating-valued, see note below this table).	P	All
OBT_END	636564618.5293884	end acquisition time in OBT	positive float	Float	OBT_END expresses the end time of data acquisition in onboard time units (double-precision floating-valued). It is only used in case the end acquisition time was contained in the TM packets. See note below this table.	O	All

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
LEVEL	'L1'	Data processing level	'L0', 'L1', 'L2', 'L3', 'LL01', 'LL02', 'ANC', 'CAL'	String	Data processing level, should be consistent with definition in Sect. 2.1.2 (Definition of Processing Levels)	P	All
ORIGIN	'Solar Orbiter SOC, ESAC'	<name of institution>	any	String	Location where the FITS file has been generated, e.g., 'SSOC' for 'Solar Orbiter Science Operations Centre', or the institute name, or ...	P	All
CREATOR	'MyFitsSoftware'	FITS creation software	Any	String	Name of the software that produced the FITS file.	P	All
VERS_SW	'2.5'	Version of SW that provided FITS file	Any	String	The version of the software that created the FITS file, as specified in CREATOR.	P	All
VERS_CAL	'2.4'	Version of calibration pack	any	String	Version number of the full package of calibration data and products applied to the data. Version should change whenever one of the included files changes.	P	L2+ (except-ionally L1)
VERSION	'03'	Incremental version number			Two-character integer, zero padded. E.g., '02'. This version number is the (only) one used in the filename.	P	All

Note on OBT:

The OnBoard Time is a double-precision floating-point number in onboard time units which is not actually decimal. It consists of a 'coarse' time and, after the decimal point, a 'fine' time for sub-seconds which is from a counter from 0-65535.

OBT = coarse + fine/65536

So for a given OBT of 637551003.4117279, expressed as a decimal in the metadata, 637551003 is the coarse part, and 4117279 is the fine -> round(0. 4117279 * 65536) = round(26982.9996544) = 26983

The value fed into the spice time conversion routine is then 637551003:26983

The archive will store the floating-point number given in the metadata so that the information is retained.

3.1.1.3 Instrument and Observation Configuration

Table 3-3 Instrument and Observation Configuration FITS Keywords

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
OBSRVTRY	'Solar Orbiter'	Satellite name	'Solar Orbiter'	String	Observatory or satellite name, in this case Solar Orbiter.	P	L1+
TELESCOP	'SOLO/EUI/FSI'	Telescope/Sensor name	"SOLO/" + $INSTRUME/ [+ $DETECTOR]	String	The full description of the telescope or sensor that took the measurement. In case of instruments with several telescope channels (different optical paths), the full chain should be specified and the channel should be added. These fields also appear in the FITS filename as 'source' and (part of) 'descriptor'.	P	L1+
INSTRUME	'EUI'	Instrument name	'EUI', 'Metis', 'PHI', 'STIX', 'SoloHI', 'EPD', 'MAG', 'SPICE', 'RPW', 'SWA'	String	This keyword should reflect the 'official' instrument acronym, i.e., how you would want it to appear in a publication. Capitalisation does not necessarily have to match the filename. Instrument names in INSTRUME should be used consistently (see valid data range). Subunits, if applicable, are specified through the keyword DETECTOR.	P	All
DETECTOR	'FSI'	Instrument subunit or sensor	*to be defined by instr teams*	String	To specify the subunit/sensor, e.g., 'FSI', 'HRI_EUV', 'HRI_LYA' for EUI, 'VL', 'UV' for Metis, ... to be	O	L1,2

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
					defined by instrument teams and used consistently.		
OBJECT	”Sun”	Type of object observed	”Sun”	String	Type of object observed	O	L1+
OBS_MODE	'EUV_CH '	Observation mode	any string (Limited list of observation modes per instrument)	String	Description of the observation mode or 'study' that has been used to acquire this image (series). This keyword/value would be used to find all data that have been acquired with similar settings (varying parameters might still apply). There should be a limited list of observation modes defined per instrument. These are also considered the building blocks ('ingredients') of the SOOP campaigns. If no OBS_ID, put a meaningful description of the mode.	P	L1,2
OBS_TYPE	‘mACv’	Encoded version of OBS_MODE	Any of the observation types as defined during planning	String	A subfield of OBS_ID that only contains an encoded version of OBS_MODE. Intended for use as a search term. This encoded ID forms part of the OBS_ID keyword value (see Table 3-4). If no OBS_ID then “none”	P	L1,2
FILTER	'<filter name>'	Description of filter used	any	String	To specify the filter in multi-filter instruments. Optional keyword, but format can be mandated based on any commonalities between instruments' proposed usage.	O	L1,2
WAVELNTH	174.0	[Angstrom] characteristic wavelength of observation	*TBD by ITs*	Float	WAVELNTH contains the characteristic wavelength at which the observation was taken, in angstroms. It is optional as it might	O	L1,2

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
					not be appropriate to all data products. It applies to instruments or sensors using a narrow bandpass. For EUV imagers, this keyword typically identifies the most prominent emission line in the bandpass.		
WAVEMIN	170.1	[Angstrom] Min wavelength where response > 0.05 of max	*TBD by ITs*	Float	The shortest wavelength at which the net (approximate) response function becomes 0.05 times the maximum response. Included for SVO-like search purposes, to identify useful observations with contributions from unexpected lines. If the response function is not known out to the wavelength(s) where this happens, the relevant min/max value(s) should be set to the limits of the area in which the response function is known/estimated. Only applies to bandpass instruments (both narrow-band and wide-band) and spectrometers.	O	L1,2
WAVEMAX	178.2	[Angstrom] Max wavelength where response > 0.05 of max	*TBD by ITs*	Float	The longest wavelength at which the net (approximate) response function becomes 0.05 times the maximum response. Included for SVO-like search purposes, to identify useful observations with contributions from unexpected lines. If the response function is not known out to the wavelength(s) where this happens, the relevant min/max value(s) should be set to	O	L1,2

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
					the limits of the area in which the response function is known/estimated. Only applies to bandpass instruments (both narrow-band and wide-band) and spectrometers.		
WAVEBAND	'ENERGY_L'	Bandpass description	any string	String	Description of the wavelength band. Typically used for wider bandpasses containing several significant lines. Recommended to be used for all instruments where appropriate, simply as human-readable description of the bandpass for people not familiar with the spectral lines.	O	L1,2
XPOSURE	10.0	[s] Total effective exposure time	Positive	Float	Total effective exposure time of the observation, in seconds. For observations made of multiple exposures summed together, this would be the total sum of the individual effective exposure times.	P	L1,2
NSUMEXP	10	Number of detector readouts summed together	1, 2, 3, ...	Integer	Number of images summed together to form the observation. The average effective exposure time of an individual exposure can be calculated as XPOSURE / NSUMEXP.	O	L1,2
TELAPSE	10.0	[s] Elapsed time between beginning and end of observation	Positive	Float	Total elapsed time between the beginning and end of the complete observation in seconds, including any dead times between exposures. In other words, this is the time from the start of the first exposure, to the end of the last	O	L1,2

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
					exposure. For rasters, this would be the duration of the entire raster.		
TRIGGERD	'STIXflag'	Event that triggered observation, default 'none'	Any string	String	Proposal, more discussion needed ⚠: This keyword specifies whether the acquisition(s) was/were triggered by a flag, either raised internally or communicated through Inter-instrument communication (service 20). Acquisitions that were completely commanded from ground have default value 'none'.		L1,2

3.1.1.4 Campaign Information

Table 3-4 Campaign Information FITS Keywords

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
SOOPNAME	**'L_SMALL_MRES_MCAD_ Connection-Mosaic'**	Name of the SOOP Campaign that the data belong to	SOOP names listed in [SAP]	String	SOOP(s) that this observation belongs to. If the data product can be linked to multiple SOOPs, the value of this keyword shall be a list of the corresponding SOOP names delimited by semicolon ';' within the same string. The source of this information is the SOOP plan as agreed upon by SOWG and distributed by SOC at the time of MTP; this will include the name of each applicable SOOP. If the file is not associated with a SOOP, put "none".	P	L1+
SOOPTYPE	**'LS4'**	Campaign ID(s) that the data belong to	see SOOP naming convention in [IOR-ICD]. Multiple values of SOOPTYPE are separated by ';'	String	Encoded identification of the SOOP(s) that this observation belongs to. Since each data product can be linked to multiple SOOPs (see SOOPNAME), in such cases the value of this keyword shall be a list of IDs delimited by semicolon ';' within the same string. The source of this information is the SOOP plan as agreed upon by SOWG and distributed by SOC at time of LTP; this information will include the encoded ID of each applicable SOOP. The value(s) of SOOPTYPE will also form part of the OBS_ID field (see below). If the file is not associated with a SOOP, put "none".	P	L1+

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
OBS_ID	'SEUI_060A_LS4_111_mACv_111'	Unique ID of the individual observation	Any of the observation IDs as defined during planning	String	Long-Term Planning identifier for the observation that is associated with the data acquisition and includes identifying information about: the source (instrument providing the data); the LTP plan ID; the SOOP type (consistent with the value in the keyword above) and the instance of the SOOP in this plan; and instrument's observation type and instance. The format of OBS_ID is defined in [IOR-ICD], and shall be included in the IORs delivered by the Instrument Teams to SOC. If several observations and/or multiple SOOPs are associated with the current data file, then the applicable OBS_IDs are delimited by ';'. If no OBS_ID then "none".	P	L1,2
TARGET	"on disc, disc centre"	Part of the sun covered by the FOV	"off-limb" "limb, East" "limb, West" "limb, North" "limb, South" "limb, North East" "limb, North West" "limb, South East" "limb, South West" "on disc" "on disc, disc centre"	String	Coarse, human interpretable pointing information.	O	L1+

The source for the campaign information in the table above shall be the science plan, distributed to the ITs after long-term planning in the format of JSON instrument timelines, and the associated IORs. Each SOOP campaign will be marked with a SOOP Type in the LTP plan, and this ID shall be used to label the IOR commands that will generate the contributing data [IOR-ICD]. At the time TM is received by the instrument teams, the resulting data products shall be marked with the same ID, either by acquisition time, or, if available, through a unique identifier that links the instrument's TM to commands.

3.1.1.5 Description of Data Content

Table 3-5 FITS Keywords for Description of Data Content

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
BSCALE*	1.0	ratio of physical to array value at 0 offset		Float	To avoid storing calibrated data as large floating-point numbers, a multiplicative factor BSCALE and an additive factor BZERO can be provided, to enable each data value to be represented as an integer ($N_{x,y}$). The true value of each pixel would then be $N_{x,y}$*BSCALE+BZERO. For raw data (counts), the default values of BSCALE and BZERO would be 1.000 and 0.000. More details on the use of these keywords in [FITSpaper] Sect. 4.4.2.5.	O	L1+
BZERO*	0.0	physical value for the array value 0		Float	See description BSCALE.	O	L1+
BTYPE	'Intensity'	Type of data		String	Descriptive text for the type of data stored in the array.	O	L1+
BUNIT*	'DN'	Units of physical value, after application of BSCALE, BZERO	Case sensitive. See [FITS+WCS]	String	Units of physical value, after application of BSCALE, BZERO. BUNIT for uncalibrated data or unitless values will be 'counts' or 'adu' (analog-to-digital converter units) where applicable, but also 'DN' (data number) can be used if the other values are not appropriate. See Greisen and Calabretta in [FITS+WCS] for a complete description of how to form unit strings.	P	L1+
DATAMIN*	10.0	minimum valid physical value	min value in data (except BLANK)	Float	DATAMIN/MAX refer to min and max data values within the file. These keywords are mandatory as they help scaling images/data. Note that DATAMIN and DATAMAX are required to be floating point numbers regardless of whether the data is floating point or integer (see [FITSpaper]).	P	All
DATAMAX*	3588.0	maximum valid physical value	max value in data (except BLANK)	Float	See DATAMIN description.	P	All
BLANK*	-2147483648	Value marking undefined pixels (before the	integer outside valid data range	Integer	The value specified for keyword BLANK gives the integer data value that is used to mark undefined pixels, i.e., before BZERO and BSCALE are applied. It shall **only be used in arrays with integer**	P	All

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
		application of BSCALE, BZERO)	[DATAMIN, DATAMAX]		**data**, i.e., with positive values of BITPIX. More information in [FITSpaper] Sect. 4.4.2.5. Each instrument team can define a value that fits their needs and should use it consistently. If there are no undefined pixels in the image, then the BLANK keyword is not required to be present. Blank pixels in floating point data are to be marked with the IEEE Not-a-Number (NaN) mechanism, and the BLANK keyword should be omitted.		
UCD	'phot.count; em.UV.10-50nm'	Unified Content Descriptor	See Appendix B (Sect. 5)	String	UCD keywords as defined by the IVOA, see Appendix B (Sect. 5) how to build them	O	L1+

3.1.1.6 Image Relative to Detector and Electronics

Table 3-6 FITS Keywords: Image Relative to Detector and Electronics

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
PXBEGn	1	First read-out pixel in dimension n	Positive, starting at 1	Integer	PXBEGn and PXENDn describe how subfields (in 2D images) and partial readouts in higher dimensional data cubes are handled. The first pixel that has been read out in dimension n is specified in PXBEGn, the last pixel in that dimension in PXENDn. For example, if PXBEG1=1 and PXEND1=NAXIS1, the detector has been fully read out in dimension 1. Note that PXBEG/END refer to pixels in unbinned images, in other words subfields are extracted before binning.	O	All
PXENDn	512	Last read-out pixel in dimension n	Positive, ≤ NAXISn	Integer	Value of the last pixel that has been read out in dimension n (see PXBEGn description).	O	All
NBINn	2	Data binning factor in dimension n	Positive, starting at 1	Integer	NBINn specifies the binning factor in the n^{th} dimension, i.e., the actual number of pixels binned in that direction. It refers to the state of the current file relative to the detector, and not the history of what happened onboard. Default value (no binning) is 1. If the rebinned pixels are kept as they are (not binning back), their resolution - as	O	All

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
					compared to the instrument's nominal resolution - can also be inferred from the resolution in arcsecs provided in the positional parameters CUNITi below. However, some instrument teams might need to convert rebinned images to their original (detector) resolution. This would be most appropriate at least for imagers that apply partial rebinning or different bins at different parts of the detector (e.g., Metis). They may need more specific keywords to describe in the metadata which binning was applied onboard, e.g., NBIN_OBn, NBIN_OB.		
NBIN	4	Total binning factor	Positive	Integer	The value of NBIN is the product of all NBINn values above. Note that it refers to the state of the current file relative to the detector.	O	All

3.1.1.7 Parameters Describing Onboard Processing

Table 3-7 FITS Keywords describing onboard processing

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
COMPRESS	'Lossless'	Data compression quality	Choice among 5 possible words, see description	String	A cross-instrument description of the data compression quality, referring to the compression applied onboard (not necessarily the state of the current file): "None" - no compression applied "Lossless" - lossless compression "Lossy-high quality" - lossy compression applied. Data quality is mostly unaffected. "Lossy-strong" - strong lossy compression applied. Analysis of data should take into account frequent compression artefacts. "Lossy-extreme" - extreme lossy compression applied. Data quality is visibly affected	O	All

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
COMP_RAT	4	Ratio of size uncompressed vs compressed image	Any	Float	The COMPRESS keyword is referring to the total compression applied to the data, including compression through binning. The value is calculated by the uncompressed information content (typically the size of the uncompressed image) divided by the compressed information content (the size of a compressed image). The aim of a numeric value is to have a sortable item, so that users/programs can pick the best observation that satisfies any other (search) criteria.	O	All

3.1.1.8 World Coordinate System Attitude Parameters

The below group of keywords (Table 3-8) describes the relationship between the detector and target. All SolO metadata on position and attitude should be expressed in WCS coordinate systems. Helioprojective coordinates are celestial spherical coordinates that use the Sun to define their pole and origin [WCSpaper]. As this is a spherical coordinate system a map projection is used. The most common one in solar physics is the TAN projection, which is used when a solar image is focused onto a flat focal plane, such as an APS detector. Helioprojective-Cartesian coordinates, as chosen in the standard below, are an angular coordinate system that measure latitude from disc centre towards solar North in theta_y, and longitude towards West limb in theta_x.

Calibrated Level-2 images should be corrected for distortion and rotation effects and always be aligned with solar North. Exceptions are allowed for 3D and higher dimensional data cubes for which rotation effects can better be described by the PCi_j rotation matrix only. A typical case are SPICE L2 data cubes.

Instrument pointing is affected by a number of factors, including light travel time, stellar aberration, and thermoelastic distortions. All of these depend to some extent on solar distance, and all interact together to determine the relative pointing of the instrument to Sun center as seen from the spacecraft. In at least Level 2 or above files, the WCS keywords below should represent the best effort to correct for these effects, e.g. by limb fitting, cross-correlation to full Sun images that have already been corrected, via trending based on the spacecraft ephemeris, temperature, etc., or other suitable means.

Table 3-8 World Coordinate System Attitude FITS Keywords

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
WCSAXES	4	Number of WCS axes	>= **NAXIS**	Integer	Number of WCS axes. This number can be larger than NAXIS, with the understanding that the additional axes are degenerate, i.e., have a dimensionality of 1. If present, must precede all other WCS keywords. An example use would be to provide 2D pointing information for a single slit-spectrograph exposure with NAXIS=2.	O	L2+
WCSNAME	'Helioprojective-cartesian'	Name of coordinate system		String	Name of coordinate system. This is typically 'Helioprojective-cartesian' for the typical image coordinate system which has solar north along the +Y axis.	P	L2+
CTYPE1	'HPLN-TAN'	Helioprojective longitude (Solar X)		String	Coordinate name and projection along axis 1. CTYPE1[2] describe the nature of the image axes - the projection of the solar disc. 'HPLN' is the Helioprojective westward angle, and 'HPLT' is the Helioprojective northward angle. The units (arcsec) are given in CUNIT1[2]. The projection is normally '-TAN', though other values are possible for telescopes with very large fields of view.	P	L2+
CTYPE2	'HPLT-TAN'	Helioprojective latitude (Solar Y)		String	Coordinate name and projection along axis 2.	P	L2+
CUNIT1	'arcsec'	Units along axis 1.	Case sensitive. See [FITS+WCS]	String	Units along axis 1. See Greisen and Calabretta in [FITS+WCS] for a complete description of how to form unit strings.	P	L2+
CUNIT2	'arcsec'	Units along axis 2.	Case sensitive. See [FITS+WCS]	String	Units along axis 2.	P	L2+
PC1_1	0.8660254	WCS coordinate transformation matrix		Float	*PCi_j* co-ordinate transformation matrix element used to express dependence of an intermediate world co-ordinate *i* on data dimension *j*. Can be used to describe the effect of the spacecraft roll angle, or more complex transformations. The example shown is for a rotation of 30 degrees.	P	L2+
PC1_2	-0.5000000	WCS coordinate transformation matrix		Float	Coordinate transformation matrix element.	P	L2+

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
PC2_1	0.5000000	WCS coordinate transformation matrix		Float	Coordinate transformation matrix element.	P	L2+
PC2_2	0.8660254	WCS coordinate transformation matrix		Float	Coordinate transformation matrix element.	P	L2+
CDELT1	3.17	[arcsec] Pixel scale along axis 1		Float	Pixel scale along axis 1 (in units consistent with CTYPE1). For projections that allow the pixel scale to vary with position, this is the pixel scale at the reference pixel location given by the **CRPIXi** keywords.	P	L2+
CDELT2	3.17	[arcsec] Pixel scale along axis 2		Float	Pixel scale along axis 2.	P	L2+
CROTA	30.0	[deg] Rotation angle		Float	Data rotation angle consistent with that expressed in the **PC*i_j*** matrix. Suitable mainly for 2D telescope images. Leaving out the axis number in the name allows this to co-exist in FITS headers with the **PC*i_j*** matrix. In past missions, the bulk rotation of the image about the direction normal to the image plane have been described by **CROTA1[2]**. However, we replace them by a more flexible and complete alternative which is to define the transformation matrix s_i*m_{ij} using the keywords **CDELT*i*** and **PC*i_j*** allowing for a skewed axes and fully general rotation of multi-dimensional arrays ([FITS+WCS]).	O	L2+
CRVAL1	0.0	[arcsec] Value of reference pixel along first axis.		Float	Value along first axis (X in example) at pixel location specified by **CRPIX*i*** keywords. Setting CRVAL1=0, CRVAL2=0 means Sun center, and the reference pixel must be calculated accordingly.	P	L2+
CRVAL2	0.0	[arcsec] Value of reference pixel along second axis.		Float	Value along second axis (Y in example) at pixel location specified by **CRPIX*i*** keywords.	P	L2+
CRPIX1	512.5	[pixel] Reference pixel location along axis 1		Float	Reference pixel location along axis 1. Must be consistent with the **CRVAL*i*** values. For a 1024x1024 image, setting the reference pixel to CRPIX1=512.5, CRPIX2=512.5 selects the center of the image, and the CRVAL*i* values must be calculated accordingly.	P	L2+

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
CRPIX2	512.5	[pixel] Reference pixel location along axis 2		Float	Reference pixel location along axis 2.	P	L2+
CRDER1	0.2	Estimated random error along axis 1	>= 0.0	Float	Representative average value over the data range of the random error in the coordinate values along axis 1, in the units given by **CUNIT1**.	O	L2+
CRDER2	0.2	Estimated random error along axis 2	>= 0.0	Float	Representative average value over the data range of the random error in the coordinate values along axis 2, in the units given by **CUNIT2**.	O	L2+
CSYER1	0.1	Estimated systematic error along axis 1	>= 0.0	Float	Representative average value over the data range of the systematic error in the coordinate values along axis 1, in the units given by **CUNIT1**. The total error is the root-summed-square of the random and systematic errors.	O	L2+
CSYER2	0.1	Estimated systematic error along axis 2	>= 0.0	Float	Representative average value over the data range of the systematic error in the coordinate values along axis 2, in the units given by **CUNIT2**. The total error is the root-summed-square of the random and systematic errors.	O	L2+
LONPOLE	180.0	[deg] Native longitude of the celestial pole		Float	Native longitude of the celestial pole. For azimuthal projections such as '-TAN', this defaults to 180 degrees, but may differ for other projections.	O	L2+
SPECSYS	'TOPOCENT'	Reference frame for spectral axes	'TOPOCENT', 'HELIOCEN'	String	Reference frame used for calculating the effect of the spacecraft velocity on spectral coordinates. Used in conjunction with **VELOSYS** keyword. Use 'HELIOCEN' if the spectral *values have been corrected* for the Doppler effect caused by the motion of the spacecraft relative to the Sun, or 'TOPOCENT' if the *values are uncorrected*.	O	L2+
VELOSYS	0.0	[m/s] Velocity correction applied to spectral data.		Float	If **SPECSYS**='HELIOCEN', then this keyword contains the radial velocity of the spacecraft relative to the Sun, in metres per second, that was used to correct the spectral coordinates. Positive velocity is outward from the Sun. If **SPECSYS**='TOPOCENT', then this keyword *must* be 0.0. See also the keyword **OBS_VR** below.	O	L2+

In addition, alternative coordinate systems can be specified by adding sets of the above WCS keywords with a single letter of the alphabet appended, e.g., **WCSNAMEA**, **CRPIX1A**, etc. Each letter specifies a separate coordinate system, and all keywords for that coordinate system, even if they don't vary from the default coordinate system, should be included. (Note that **SPECSYS** and **VELOSYS** actually relate to the orbital velocity of the spacecraft, rather than the spacecraft attitude, but are included here because they also can have an alternate coordinate letter appended to them, e.g., **SPECSYSA**, **VELOSYSA**.)

3.1.1.8.1 Attitude Parameters (Example for Spectral Data)

The following example shows how the definitions above *could* be extended to a three-dimensional data cube containing spectral data.

```
NAXIS = 3
NAXIS1 = 1024  / Number of pixels in spectral dimension
NAXIS2 = 10 / Number of mirror positions in raster
NAXIS3 = 800 / Number of pixels along slit
WCSNAME = 'Helioprojective-cartesian'
CTYPE1 = 'WAVE' / Wavelength
CTYPE2 = 'HPLN-TAN' / Helioprojective longitude (Solar X)
CTYPE3 = 'HPLT-TAN' / Helioprojective latitude (Solar Y)
CUNIT1 = 'Angstrom ' / Units along spectral dimension
CUNIT2 = 'arcsec ' / Units along second axis (Solar X)
CUNIT3 = 'arcsec ' / Units along third axis (Solar Y)
CDELT1 = 0.1 / [Angstrom] Wavelength scale
CDELT2 = 4.0 / [arcsec] Scan step size
CDELT3 = 1.1 / [arcsec] Slit pixel size, resolution along slit
CRVAL1 = 300. / [Angstrom] Reference wavelength
CRVAL2 = 0.0 / [arcsec] Solar X coordinate of reference pixel
CRVAL3 = 30.0 / [arcsec] Solar Y coordinate of reference pixel
CRPIX1 = 512.5 / [pixel] Pixel coord. of reference wavelength
CRPIX2 = 1.0 / [pixel] Reference pixel is first exposure in raster
CRPIX3 = 400.5 / [pixel] Reference pixel is slit center
PC1_1 = 1.0 / No wavelength transformation
```

```
PC2_2 = 1.0 /  No rotation of SPICE scan direction wrt solar north
PC3_3 = 1.0 /  No rotation of SPICE scan direction wrt solar north
```

3.1.1.9 Solar Ephemeris Data Keywords

These are based on good practice from SECCHI and SWAP (note that example values may not be consistent with each other).
All ephemeris values are to be calculated at the time DATE-AVG.

With the sole exception of CRLN_OBS as described below, the positional ephemeris keywords listed below should not include any corrections for light travel time or stellar aberration. This is because files should be consistent amongst all Solar Orbiter instruments (including in situ) first and foremost. The reason for the exception is to be consistent with other missions (e.g. SDO) and observatories.

There is a certain ambiguity in the definition of CRLN_OBS in that it can either refer to the position of the spacecraft or the Carrington longitude of the Sun's central meridian as seen from the spacecraft at the time of observation. Because the Carrington system rotates at a sidereal rate of once every 25.38 days, these two definitions differ by a small amount due to light travel time, and which is significant for Solar Orbiter's high resolution imagery. It has been decided that the value of CRLN_OBS in Solar Orbiter FITS headers should be corrected for this effect. The conversion from ephemeris to corrected Carrington longitude values is given by the following relationship:

$$\mathrm{CRLN\_OBS}_{\mathrm{corrected}} = \mathrm{CRLN\_OBS}_{\mathrm{ephemeris}} + \rho \times (D_{\mathrm{Sun}} - R_{\mathrm{Sun}}) / c$$

Where ρ is the Carrington sidereal rate (14.18440 deg/day), D_{Sun} is the solar distance of the spacecraft from Sun center, R_{Sun} is the solar radius, and c is the speed of light. An example of this correction is given in the SolarSoft routine get_sunspice_lonlat.pro through the /LT_CARR keyword. If this results in a value greater than 360°, then 360° should be subtracted from the result, and the value of CAR_ROT should also be decremented by one to reflect that it's part of the previous rotation.

Table 3-9 Solar Ephemeris Data FITS Keywords

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
RSUN_ARC	952.565823288377	[arcsec] Apparent photospheric solar radius	Positive	Float	Apparent photospheric solar radius in arc seconds.	P	L2+
RSUN_REF	6.957E+08	[m] Assumed physical solar radius	Positive	Float	Assumed photospheric solar radius, in metres. The intended purpose of this keyword is to document what assumption was made in determining the instrument plate scale, so that it can be backed out for comparison with measurements assuming a different solar radius. SOC will use the value in IAU resolution B3 (2015) [IAU15]	O	L2+
SOLAR_B0	-4.18794601814401	[deg] s/c tilt of solar North pole		Float	The tilt angle of the solar North pole toward (positive) or away from (negative) the spacecraft, in degrees. Equals **HGLT_OBS**, **CRLT_OBS**.	O	L2+
SOLAR_P0	16.242648	[deg] s/c celestial North to solar North angle		Float	The apparent angle in the sky, as seen from the spacecraft location, between celestial north and solar north.	O	L2+
SOLAR_EP	5.92376011723103	[deg] s/c ecliptic North to solar North angle		Float	The apparent angle in the sky, as seen from the spacecraft location, between ecliptic north and solar north.	O	L2+
CAR_ROT	2158	Carrington rotation number		Integer	Carrington rotation number associated with pixel given by CRPIX*i*.	P	L2+
HGLT_OBS	-4.18794601814401	[deg] s/c heliographic latitude (B0 angle)		Float	Stonyhurst heliographic latitude of the spacecraft, in degrees.	P	L2+
HGLN_OBS	0.00131206248030214	[deg] s/c heliographic longitude		Float	Stonyhurst heliographic longitude of the spacecraft, in degrees.	P	L2+

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
CRLT_OBS	-4.18794601814401	[deg] s/c Carrington latitude (B0 angle)		Float	Carrington heliographic latitude of the spacecraft, in degrees. Because of the way the two coordinate systems are defined, HGLT_OBS and CRLT_OBS are always the same. However, both should appear in the FITS header.	P	L2+
CRLN_OBS	25.6633839152	[deg] s/c Carrington longitude (L0 angle)		Float	Carrington heliographic longitude of the spacecraft, in degrees, corrected for light travel time (see above).	P	L2+
DSUN_OBS	150707994180.746	[m] s/c distance from Sun	Positive	Float	Distance of the spacecraft from the center of the Sun, in metres.	P	L2+
DSUN_AU	1.0074207171235	[AU] s/c distance from Sun	Positive	Float	Distance of the spacecraft from the center of the Sun, in AU.	O	L2+
HEEX_OBS	150707994015.506	[m] s/c Heliocentric Earth Ecliptic X		Float	X axis position of the spacecraft in Heliocentric Earth Ecliptic coordinates, in metres.	P	L2+
HEEY_OBS	4059466.73665196	[m] s/c Heliocentric Earth Ecliptic Y		Float	Y axis position of the spacecraft in Heliocentric Earth Ecliptic coordinates, in metres.	P	L2+
HEEZ_OBS	5772938.06008995	[m] s/c Heliocentric Earth Ecliptic Z		Float	Z axis position of the spacecraft in Heliocentric Earth Ecliptic coordinates, in metres.	P	L2+
HCIX_OBS	109043378382.	[m] s/c Heliocentric Inertial X		Float	X axis position of the spacecraft in Heliocentric Inertial coordinates, in metres.	P	L2+
HCIY_OBS	94332806329.1	[m] s/c Heliocentric Inertial Y		Float	Y axis position of the spacecraft in Heliocentric Inertial coordinates, in metres.	P	L2+
HCIZ_OBS	-12321790987.1	[m] s/c Heliocentric Inertial Z		Float	Z axis position of the spacecraft in Heliocentric Inertial coordinates, in metres.	P	L2+

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
HCIX_VOB	-24982.158	[m/s] s/c Heliocentric Inertial X Velocity		Float	Spacecraft velocity along the X axis in Heliocentric Inertial coordinates, in metres per second. Keywords for velocities in other coordinate systems can also be formed by substituting _VOB for _OBS, but will not be inertial velocities, and are therefore optional.	P	L2+
HCIY_VOB	-17026.140	[m/s] s/c Heliocentric Inertial Y Velocity		Float	Spacecraft velocity along the Y axis in Heliocentric Inertial coordinates, in metres per second.	P	L2+
HCIZ_VOB	2231.6757	[m/s] s/c Heliocentric Inertial Z Velocity		Float	Spacecraft velocity along the Z axis in Heliocentric Inertial coordinates, in metres per second.	P	L2+
HAEX_OBS	-65400700049.4	[m] s/c Heliocentric Aries Ecliptic X		Float	X axis position of the spacecraft in Heliocentric Aries Ecliptic coordinates, in metres.	P	L2+
HAEY_OBS	129087613712.	[m] s/c Heliocentric Aries Ecliptic Y		Float	Y axis position of the spacecraft in Heliocentric Aries Ecliptic coordinates, in metres.	P	L2+
HAEZ_OBS	-315707554.534	[m] s/c Heliocentric Aries Ecliptic Z		Float	Z axis position of the spacecraft in Heliocentric Aries Ecliptic coordinates, in metres.	P	L2+
HEQX_OBS	-139113055426.	[m] s/c Heliocentric Earth Equatorial X		Float	X axis position of the spacecraft in Heliocentric Earth Equatorial coordinates, in metres.	P	L2+
HEQY_OBS	37903753496.0	[m] s/c Heliocentric Earth Equatorial Y		Float	Y axis position of the spacecraft in Heliocentric Earth Equatorial coordinates, in metres.	P	L2+
HEQZ_OBS	-12321790987.1	[m] s/c Heliocentric Earth Equatorial Z		Float	Z axis position of the spacecraft in Heliocentric Earth Equatorial coordinates, in metres.	P	L2+

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
GSEX_OBS	632918.814814495	[m] s/c Geocentric Solar Ecliptic X		Float	X axis position of the spacecraft in Geocentric Solar Ecliptic coordinates, in metres. *Note: There's some discussion in the WCS group on whether geocentric coordinates are useful for Solar Orbiter.*	P	L2+
GSEY_OBS	-4059466.73664814	[m] s/c Geocentric Solar Ecliptic Y		Float	Y axis position of the spacecraft in Geocentric Solar Ecliptic coordinates, in metres.	P	L2+
GSEZ_OBS	5772938.06009963	[m] s/c Geocentric Solar Ecliptic Z		Float	Z axis position of the spacecraft in Geocentric Solar Ecliptic coordinates, in metres.	P	L2+
OBS_VR	-165.20536	[m/s] Radial velocity of spacecraft relative to Sun		Float	Radial velocity of the spacecraft relative to the Sun, in metres per second. Positive velocity is outward from the Sun. If the **SPECSYS** keyword is set to 'TOPOCENT', then this velocity can be used to correct the spectral coordinates for the Doppler effect.	P	L2+
EAR_TDEL	349.06169	[s] Time(Sun to Earth) - Time(Sun to S/C)		Float	Difference in light travel time between the Sun to Earth (heliocentre to geocentre) and the light travel time from the Sun to the spacecraft (heliocentre to s/c), in seconds. Positive if the spacecraft-Sun distance is smaller than the Sun-Earth distance.	P	L2+
SUN_TIME	149.59788	[s] Time(Sun to S/C)	Positive	Float	Light travel time from the Sun (heliocentre) to the spacecraft.	P	L2+
DATE_EAR	'2020-01-22T14:05:49.062'	[UTC] Start time of observation, corrected to Earth	any date+'T'+time within mission duration	String	Start time of observation corrected for the difference in light travel between the Sun to Earth (heliocentre to geocentre) and the light travel time from the Sun to the spacecraft (heliocentre to s/c), in UTC, in ISO-8601 format. In other words, this is DATE-BEG *plus* EAR_TDEL.	P	L2+
DATE_SUN	'2020-01-22T13:57:30.402'	[UTC] Start time of observation, corrected to Sun	any date+'T'+time within mission duration	String	Start time of observation corrected for the difference in light travel from the Sun to the spacecraft (heliocentre to s/c), in UTC, in ISO-8601 format. In other words, this is DATE-BEG *minus* SUN_TIME.	P	L2+

3.1.1.10 Parameters Closing Metadata Description

Table 3-10 FITS Keywords Closing Metadata Description

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
INFO_URL	'$<DOI instrument paper or URL of information page'	Link to additional information	string	String	Link to more information on the instrument data. Could contain Bibcode or DOI of instrument paper, or URL of instrument page with complete and updated information on the data and how to use it.	O	L1+
COMMENT*	'$<comment text>$'				This keyword may be used to supply any comments regarding the FITS file.	O	All
CHECKSUM	'VVdCWTZ9VTbCVTZ9'	HDU checksum updated at <date/time>	character string	String	The value field of the CHECKSUM keyword shall consist of an ASCII character string whose value forces the 32-bit 1's complement checksum accumulated over the entire FITS HDU to equal negative 0. Documentation on how to generate the checksum in [FITS Checksum].	P	All
DATASUM	'2189405276'	data unit checksum updated at <date/time>	character string	String	The value field of the DATASUM keyword shall consist of a character string containing the unsigned integer value of the 32-bit 1's complement checksum of the data records in the HDU (i.e., excluding the header records). Documentation on how to generate the data checksum in [FITS Checksum].	P	All
HISTORY*	'<software name>'			String	This keyword should be used to describe the history of steps and procedures associated with the processing of the associated data.	P	All
END					This keyword has no associated value. Bytes 9 through 80 shall be filled with ASCII spaces (decimal 32 or hexadecimal 20). The END keyword marks the logical end of the header and must occur in the last 2880-byte FITS block of the header.	M	All

3.1.2 FITS Extensions

Besides the primary header and data unit (HDU), a FITS file may contain one or more HDU extensions. These extensions are formatted in the same manner as the primary HDU, with an ASCII text header followed by a data unit. The following keywords apply to FITS extensions in general.

Table 3-11 General FITS Extension Keywords

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
XTENSION	'IMAGE '	Extension type	'IMAGE ', 'TABLE ', 'BINTABLE'	String	Name of the type of extension being used. Must be one of the three currently defined conforming extensions, 'IMAGE ', 'TABLE ', or 'BINTABLE', must be the first keyword in the header, and must be padded out with spaces to be at least 8 characters long. *Cannot* appear in the primary HDU.	M	All
BITPIX	8	Number of bits per data unit.		Integer	Number of bits per data unit. Value may depend on the type of extension.	M	All
NAXIS	2	Number of axes in data cube		Integer	Number of data axes. Value may depend on the type of extension.	M	All
NAXISn	128	Number of data units along axis n.		Integer	Number of data units along axis n. This has different meanings in different extension types.	M	All
PCOUNT	0	Parameter count			Value depends on the type of extension.	M	All
GCOUNT	1	Group count			Value depends on the type of extension.	M	All
EXTNAME		Extension name		String	Name of the extension. May also appear in the primary HDU.	P	All

Depending on the type of extension, some additional keywords may be mandatory, as discussed below.

3.1.2.1 Types of extensions

3.1.2.1.1 IMAGE extensions

By far the simplest type of extension is the IMAGE extension. This is formatted exactly the same as the primary HDU, with the exception that the **XTENSION** keyword replaces the **SIMPLE** keyword, and the header will contain the keyword/value combinations PCOUNT=0, GCOUNT=1.

Otherwise, all the keywords defined for the primary header are equally applicable to IMAGE extensions, with the same requirements on which keywords are mandatory or optional (except EXTEND, see [FITSpaper] Section 4.4.2.1 p13). An example header for an IMAGE extension would be:

```
XTENSION= 'IMAGE   ' / Extension type
BITPIX  = 16         / Number of bits per data pixel
NAXIS   = 2          / Number of data axes
NAXIS1  = 512        / Number of pixels along axis 1
NAXIS2  = 512        / Number of pixels along axis 2
PCOUNT  = 0          / Parameter count
GCOUNT  = 1          / Group count
EXTNAME = 'Helium II 304 Angstroms'
    ...
END
```

3.1.2.1.2 Binary table (BINTABLE) extensions

A much more complicated, but very powerful, type of extension is the binary table extension, expressed by setting XTENSION='BINTABLE'. A binary table is organized as rows and columns. Each column has a data type, which may be multidimensional. The following primary keywords are used to set up the binary table.

Table 3-12 FITS Binary Extension Table Basic Keywords

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
XTENSION	'BINTABLE'	Extension type	'BINTABLE'	String	Specifies that what follows is a binary table.	M	All
BITPIX	8	Number of bits per data unit.		Byte	This is always 8 for binary tables. The actual data types within the table are specified separately.	M	All
NAXIS	2	Number of axes in data cube		Byte	This is always 2 for binary tables, to signify that the table is organized into columns and rows.	M	All
NAXIS1	128	Number of bytes in each data row.	Positive	Integer	Number of bytes in each data row, summed from all the columns in the table.	M	All
NAXIS2	10	Number of rows in the table	Positive	Integer	Number of rows in the table.	M	All

Keyword	Example Value	Suggested FITS Comment	Valid Data Range	Value Type	Description	M-Type	Level
PCOUNT	0	Number of bytes following table			Number of bytes that follow the table in the supplemental data area called the heap. This is to support the variable length arrays option.	M	All
GCOUNT	1	Group count			This is always 1 for binary tables.	M	All
TFIELDS	5	Number of table columns	Positive	String	Number of columns in the table.	M	All
TFORMn	'20I'	Data format		String	String describing the format of the data in column n, where n runs from 1 to **TFIELDS**. The example '20I' represents a vector of 20 16-bit integer values. See [FITSpaper] for a complete description of how **TFORMn** keywords are formed.	M	All
TTYPEn	'Intensity'	Data type		String	Name of the data in column n.	P	All
TUNITn	'DN/s'	Data units		String	Units of the data in column n. See description for **BUNIT** keyword.	P	All
TDIMn	'(5,4)'	Data dimensions		String	String specifying the dimensions of multi-dimensional data. Only mandatory for multi-dimensional data.	O	All

A number of the keywords defined for primary header have equivalents defined for use in binary table columns. These equivalents are given in the following table, where ***n*** represents the binary table column number, ***i*** and ***j*** represent axis numbers (from 1-9), and ***a*** represents an optional letter from A to Z signifying an alternate coordinate system.

Table 3-13 FITS Primary Header Equivalent Keywords for Binary Extensions

Keyword Description	Primary Array	BINTABLE vector	
		Primary	Secondary
Coordinate dimensionality	**WCSAXESa**	**WCAXna**	
Axis type	**CTYPEia**	**iCTYPn**	**iCTYna**
Axis units	**CUNITia**	**iCUNIn**	**iCUNna**
Reference value	**CRVALia**	**iCRVLn**	**iCRVna**
Coordinate increment	**CDELTia**	**iCDLTn**	**iCDEna**
Reference point	**CRPIXja**	**jCRPXn**	**jCRPna**
Transformation matrix	**PCi_ja**	**ijPCna**	
Transformation matrix	**CDi_ja**	**ijCDna**	

Keyword Description	Primary Array	BINTABLE vector	
		Primary	Secondary
Coordinate name	**WCSNAMEa**	**WCSNna**	
Random error	**CRDERia**	**iCRDna**	
Systematic error	**CSYERia**	**iCSYna**	
Coordinate native longitude	**LONPOLEa**	**LONPna**	
Date/time of observation	**DATE-OBS**	**DOBSn**	
Average date/time of observation	**DATE-AVG**	**DAVGn**	
Scaling parameter	**BSCALE**	**TSCALn**	
Zero offset	**BZERO**	**TZEROn**	
Null value for integers	**BLANK**	**TNULLn**	

Keywords which apply to the HDU as a whole, and which do not vary between columns, are put in the header in the normal fashion. For example, if all the data has the same average observation time, then **DATE-AVG** is placed in the header instead of the individual **DAVGn** for every column.

The FITS Green Bank convention allows keyword values to be stored in a column of the binary table. Such values then pertain all the columns in the table, but can vary by row. This is done by setting the value of **TTYPEn** to the name of the keyword. For example, if column 10 in the table had TTYPE10 = 'DATE-BEG', then the contents of this column would have the values of the beginning date/time of the observation for each row in the table. The datatype of the column would be appropriate for the keyword that it is storing. Thus, TTYPE10 = 'DATE-BEG' would be stored as an ASCII string column, while TTYPE11 = 'CRVAL1' would be stored as a floating pointing scalar column. More information about the FITS Green Bank convention is available at the Registry of FITS Conventions at http://fits.gsfc.nasa.gov/fits_registry.html.

3.1.3 Compressed FITS

Compressed FITS files are described in FITS Standard v4.0 [FITSpaper] Section 10. There are no new requirements for metadata in a compressed FITS image file than for an uncompressed FITS image. The internal structure of the file is different, but this is handled by the archive and is transparent to the user.

3.2 CDF Metadata Standard

This section presents the data and metadata conventions for the Solar Orbiter CDF Level 1 (engineering data) and Level 2 (calibrated science data) data sets. The CDF format shall be used for those time series data files produced by the in situ instruments. Metadata conventions for the CDF data sets shall be as much as possible compliant with the International Solar Terrestrial Physics (ISTP) guidelines [ISTPstandard].

Section 3.2.1 describes the CDF variable recommendations. Sections 3.2.2 and 3.2.3 give respectively the lists of global and variable attributes to be used. In the global/variable attribute lists, grouped by metadata type, we distinguish 3 types of keywords:

- M: Mandatory keywords to comply with the ISTP guidelines.
- P: Keywords we propose to be mandatory in the Solar Orbiter CDF files.
- O: Keywords in *italic* are optional and can be defined if applicable to the instrument in question.

Any additional variables and attributes specific to instrument CDF data sets can be implemented as long as they are referenced in the dedicated dataset description documents.

3.2.1 CDF variables

In a CDF format file, each data parameter (e.g., time, flux, etc.) is represented by a given CDF variable. This CDF variable is identified by its name and described by a list of given variable attributes.

There are 3 types of CDF variables:

- "Data", which corresponds to the variables of primary importance (e.g, density, magnetic fields, flux).
- "Support_data", which corresponds to the variables of secondary importance (e.g., time, frequency, energy_bands).
- "Metadata", which provides labels for "data" and "support_data" types (e.g., "Bx, By, Bz" labels for a magnetic field components "data" variable).

3.2.1.1 General conventions

The general conventions for the CDF variables for Solar Orbiter are provided in the following list:

- CDF variable description and naming conventions shall be compliant with the ISTP guidelines. In addition, CDF variable names shall contain capital letters only and shall not exceed 63 characters.
- CDF variables shall be described using the appropriate variable attributes (see Sect. 3.2.3 on "Variable attributes" below).
- CDF data files shall contain at least the expected variables listed in the Sect. 3.2.1.5.

3.2.1.2 Time conventions

The time conventions for the CDF variables are as follows:

- According to the ISTP guidelines, all of the CDF files shall contain at least one primary CDF variable called “EPOCH”, which shall be the time reference variable. This variable shall be a CDF_TIME_TT2000 data type, and it shall be the first to be defined in the skeleton files.
- Data sets including multiple temporal resolutions shall use “EPOCH_i” (i=1,N) variables to define the multiple time samples after the first, which should still be “EPOCH”.
- Level 1 and Low Latency data files shall provide the onboard S/C time values in a CDF variable "SCET". See note below.
- In addition to “EPOCH” and "SCET", level 1 and low latency data files should also contain a CDF variable named “ACQUISITION_TIME”. This variable can be used to store the CCSDS format time values as written in the TM packet data.
- Any other time variables derived from “EPOCH” (e.g., milliseconds of day, Julian days, etc.) can be provided using the CDF variable conventions.

It should be noted that the CDF_TIME_TT2000 data type is defined as an 8-byte signed integer with a fixed Time_Base=J2000 (Julian date 2451545.0 TT or 2000 January 1, 12h TT), Resolution=nanoseconds, Time_Scale=Terrestrial Time (TT), Units=nanoseconds, Reference_Position=Rotating Earth Geoid. The recommended fill value for CDF_TIME_TT2000 data is -9223372036854775807.

Given a current list of leap seconds, conversion between TT and UTC is straightforward:

TT = TAI + 32.184s or
TT = UTC + deltaAT + 32.184s

Where deltaAT is the sum of the leap seconds since 1960; for example, for 2009, deltaAT = 34s).

Note on SCET:

The SCET is a double-precision floating-point number in spacecraft time units which is not actually decimal. It consists of a ‘coarse’ time and, after the decimal point, a ‘fine’ time for sub-seconds which is from a counter from 0-65535.

SCET = coarse + fine/65536

So for a given SCET of 637551003.4117279, expressed as a decimal in the metadata, 637551003 is the coarse part, and 4117279 is the fine -> round(0. 4117279 * 65536) = round(26982.9996544) = 26983

The value fed into the spice time conversion routine is then 637551003:26983

3.2.1.3 Coordinate system conventions

The coordinate system conventions for the CDF variables are as follows:

- Level 2 vector and tensor time series should always be presented in a scientific coordinate system. Spacecraft-centric RTN coordinates in a Cartesian representation are the preferred choice for Solar Orbiter.
- Further variables containing the same data in spacecraft or instrument coordinates are optional. In the event that instrument rather than spacecraft coordinates are included, the rotation matrix from instrument to spacecraft coordinates should also be included in the file.
- Level 1 data should be presented in instrument coordinates, together with the rotation matrix to spacecraft coordinates.
- Coordinate systems will be described using four variable attributes:
    1. COORDINATE_SYSTEM which gives the name of the coordinate system;
    2. FRAME_VELOCITY which can take either the value 'Observatory' where no corrections have been applied to the data or 'Inertial' where quantities (e.g., electric field or plasma flow velocity) have been corrected for spacecraft motion relative to an inertial frame (HCI);
    3. FRAME_ORIGIN which gives the origin of the reference frame where this is not implicit in the value of COORDINATE_SYSTEM;
    4. REPRESENTATION_i which gives the representation (['x','y','z'] for Cartesian; ['r','p','t'] for spherical polar; ['r','p','z'] for cylindrical polar) of the ith dimension of the variable.

3.2.1.4 Data quality

CDF files shall include the QUALITY_FLAG and QUALITY_BITMASK variables to inform on data quality. They are both time dependent parameters and can apply to one or several science parameters.

QUALITY_FLAG shall be a CDF_UINT1 flag providing a human readable high-level information about the quality with the following values:

Table 3-14 CDF Quality Flag allowed values.

Quality	Meaning
0	Bad data
1	Known problems, use at your own risk
2	Survey data, possibly not publication-quality
3	Good for publication, subject to PI approval
4	Excellent data which has received special treatment

QUALITY_BITMASK shall be a CDF_UINT2 flag providing a computer readable quality information. The values and meanings of QUALITY_BITMASK can differ from instrument to another and shall be reported in the instrument's dataset description document.

3.2.1.5 Expected variables

The following table provides a summary list of the variables expected to be found in the Solar Orbiter CDF data files. The "LEVEL" column indicates the processing level(s) for which the variable is mandatory or optional.

Table 3-15 Expected Variables in Solar Orbiter CDF Files

NAME	DESCRIPTION	TYPE	LEVEL	DATA TYPE	COMMENT
EPOCH	Primary time variable as defined in the section 3.2.1.2	M	All	CDF_TIME_TT2000	See http://spdf.gsfc.nasa.gov/istp_guide/variables.html#Epoch.
EPOCH_i	i-th time variable as defined in the section 3.2.1.2	M	All	CDF_TIME_TT2000	Only mandatory for multiple temporal resolutions.
SCET	Onboard S/C time	P	L1	CDF_REAL8	See Section 3.2.1.2 for notes on SCET
ACQUISITION_TIME	Time of acquisition as returned in the instrument packet data	O	L1	TBD	Not necessarily the same as SCET
JULIAN_DAY	Time in Julian days	O	L2, L3	CDF_REAL8	
SYNCHRO_FLAG	Flag to check instrument time synchronisation	O	L1	CDF_UINT1	e.g., Check time synchronisation between the RPW DPU time and sub-systems times.
QUALITY_FLAG	Human readable high-level parameter	P	All	CDF_UINT1	Can apply to one or several science parameters. Should be record-varying. Applies to variables that depend on EPOCH.

NAME	DESCRIPTION	TYPE	LEVEL	DATA TYPE	COMMENT
QUALITY_FLAG_i	i-th quality flag variable	P	All	CDF_UINT1	Can apply to one or several science parameters. Should be record-varying. Applies to variables that depend on EPOCH_i. Only mandatory when "EPOCH_i" is defined.
QUALITY_BITMASK	Computer readable quality parameter	P	All	CDF_UINT2	Can apply to one or several science parameters
QUALITY_BITMASK_i	i-th quality bitmask variable	P	All	CDF_UNIT2	Can apply to one or several science parameters. Should be record-varying. Applies to variables that depend on EPOCH_i. Only mandatory when "EPOCH_i" is defined.
POST_GAP_FLAG	Flag that indicates data gap in the records.	O	All	CDF_UINT1	See http://spdf.gsfc.nasa.gov/istp_guide/variables.html#Post%20Gap%20Flag
INTERPOL_FLAG	Flag that indicates if the current record is real or interpolated	O	All	CDF_UINT1	
SBM_FLAG	Flag that indicated if a SBM mode is currently on/off.	O	All	CDF_UINT1	

3.2.2 Global attributes

Global attributes are used to provide information about the data set as an entity. Global attributes for Solar Orbiter CDF data sets are divided into 4 categories:

- Attributes defined by the ISTP (http://spdf.gsfc.nasa.gov/sp_use_of_cdf.html)
- Attributes associated to the Space Physics Archive Search and Extract dictionary (SPASE, http://www.spase-group.org/)

- Attributes derived from the Virtual European Solar and Planetary Access dictionary (VESPA, http://voparis-europlanet-new.obspm.fr/planetary/data/epn/query/all/)
- Additional attributes specific to Solar Orbiter CDF data sets.

All of the global attributes shall be CDF_CHAR data types.

3.2.2.1 ISTP Global Attributes

The following table provides the list of ISTP global attributes that should be used for Solar Orbiter CDF data sets.

Table 3-16 ISTP Global Attributes for Solar Orbiter Files

NAME	DESCRIPTION	TYPE	DEFAULT VALUE	COMMENT
Project	The name of the project.	M	“SOLO>Solar Orbiter”	Use the default.
Source_name	The mission or investigation that contains the sensors. It shall use the ISTP format "PREFIX>Suffix" (e.g., "SOLO>Solar Orbiter"), where the value of the prefix shall correspond to the "source" field in the file naming convention.	M	"SOLO>Solar Orbiter"	Use the default.
Discipline	The science discipline and subdiscipline.	M	"Space Physics>Interplanetary Studies"	Use the default.
Data_type	ISTP defined exchangeable data product type, e.g., L2>Level 2 Data	M		For Solar Orbiter This will be the same as the processing level of the data. See also http://spdf.gsfc.nasa.gov/istp_guide/gattributes.html#Data_type
Descriptor	The descriptor plus full words to describe the descriptor. This shall be the same as that given in the filename, plus the relevant part of the logical_source_description, e.g.,	M		Note that the descriptor field in the file name is in fact the combination of the instrument [ISTP descriptor attribute] and the Solar Orbiter data_product attribute. Hence, in the example given, the data_product would be

NAME	DESCRIPTION	TYPE	DEFAULT VALUE	COMMENT
	EPD-EPT-ASUN-BURST-ELE-CLOSE>Energetic Particle Detector, Electron Proton Telescope, Anti-Sun direction, Burst, Electrons, Close mode			BURST-ELE-CLOSE>Burst Electrons Close Mode
Instrument	Acronym and name of instrument and sensor, e.g., SWA-EAS>Solar Wind Analyser Electron Analyser System	M		
Data_version	This attribute identifies the version of a particular CDF data file for a given generation date.	M		The value of the version shall be consistent with the "version" field in the filename and shall be an integer increment starting at 01.
Software_version	The version of the software that generated the CDF.	P		
Skeleton_version	The skeleton file version number.	O		It shall be an integer increment starting at 01.
PI_name	Name of the PI.	M		e.g., "Helen Middleton
PI_affiliation	PI affiliation.	M		e.g., "ESAC (ESA), Madrid".
TEXT	Description of the experiment.	M		This shall be used to describe the instrument and to provide references. The first entry shall be reserved for the description, followed by as many entries as references. Reference entries shall be sorted by decreasing importance.
Instrument_type	The ISTP defined instrument type. Multi-valued.	M		See http://spdf.gsfc.nasa.gov/istp_guide/gattributes.html#Instrument_type
Mission_group	The assigned name of the mission or project.	M	"Solar Orbiter"	Use the default.
Logical_source	Source_name, level, and full descriptor, i.e., filename up to date e.g., solo_L1_mag-obs-normal	M		See http://spdf.gsfc.nasa.gov/istp_guide/gattributes.html#Logical_source
Logical_file_id	The name of the CDF file	M		See http://spdf.gsfc.nasa.gov/istp_guide/gattributes.html#Logical_file_id

NAME	DESCRIPTION	TYPE	DEFAULT VALUE	COMMENT
Logical_source_description	Full words associated with the Logical_source, e.g., Solar Orbiter, Level 2 Data, Energetic Particle Detector, Electron Proton Telescope, Anti-Sun direction, Burst, Electrons, Close mode	M		See http://spdf.gsfc.nasa.gov/istp_guide/gattributes.html#Logical_source_description
File_naming_convention	Description of the file naming convention: source_datatype_descriptor_ yyyyMMddThhmnss	O	source_datatype_descriptor _ yyyyMMddThhmnss	It shall be consistent with the file naming convention laid out in Sect. 2.1 (Data Formats and File Names),
Rules_of_use	Citability and PI access restrictions. This may point to a World Wide Web page specifying the rules of use.	M		
Generated_by	The generating data centre/group.	M		The name of the Institute that produced the data file.
Generation_date	Date stamp for the creation of the file.	M		This shall be the data file creation date in ISO8601 format ("YYYY-MM-DDTHH:MN:SS")
Acknowledgement	Text string at PI disposal allowing for information on expected acknowledgment if data is citable.	M		
MODS	History of modifications made to the CDF data set.	P		A new entry shall be added each time a modification is made to the CDF data set. Each entry has to specify the date of the modifications, the person responsible and a summary of the changes.
LINK_TEXT	Text describing on-line data available at PI or Co-I web sites.	O		TBD
LINK_TITLE	The title of the web site holding on-line data available at PI or Co- I web sites.	O		TBD
HTTP_LINK	The URL for the PI or Co-I web site holding on-line data.	O		If such a site exists.
TEXT_supplement_1	An attribute that can be used for providing additional information about dataset.	O		
Parents	Name(s) of the CDF file(s) that were processed to become this file, for files of derived and	P		

NAME	DESCRIPTION	TYPE	DEFAULT VALUE	COMMENT
	merged data sets. Subsequent entry values are used for multiple parents. The syntax for a CDF parent would be e.g., "CDF>logical_file_id".			

3.2.2.2 SPASE attributes

The following table provides the list of SPASE (http://www.spase-group.org/) global attributes for Solar Orbiter CDF data sets.

Table 3-17 SPASE attributes for Solar Orbiter CDF Files

NAME	DESCRIPTION	TYPE	COMMENT
CAVEATS	Information which may be important in the avoidance of the misuse of the resource, for instance the assumptions or limitations on data processing modelings or inversions.	O	
spase_DatasetResourceID	The SPASE ResourceID assigned of the NumericalData resource the data file is part of.	O	
spase_DatasetResource	The SPASE XML description of the dataset that corresponds to the SPASE ResourceID.	O	
spase_GranuleResourceID	The Granule ResourceID assigned to the data file.	O	
spase_GranuleResource	The SPASE XML description of the dataset that corresponds to the SPASE Granule ResourceID	O	

3.2.2.3 VESPA attributes

The following table provides the list of the Solar Orbiter CDF global attributes derived from the VESPA keywords.

Table 3-18 VESPA Attributes for Solar Orbiter CDF Files

NAME	DESCRIPTION	TYPE	COMMENT
TARGET_NAME	Name of the target	P	e.g., 'Sun'
TARGET_CLASS	Class of the target	P	e.g., 'Star'
TARGET_REGION	Region where the target in the observed	P	e.g., 'Corona'

NAME	DESCRIPTION	TYPE	COMMENT
ACCESS_FORMAT	Format of the file	O	e.g., 'CDF' or 'CDF V3.6
DATAPRODUCT_TYPE	High level scientific organization of the data	O	
TIME_MIN	The date and time of the beginning of the first acquisition for the data contained in the file, in ISO format	P	
TIME_MAX	The date and time of the end of the last acquisition for the data contained in the file, in ISO format	P	
REFERENCE	Bibcode, DOI or URL	O	

More details about the VESPA keywords can be found in the appendix VESPA keywords for Unified Solar Orbiter Catalog (USOC)

3.2.2.4 Additional attributes

The following table provides the list of additional global attributes for Solar Orbiter CDF data sets.

Table 3-19 Additional Global Attributes for Solar Orbiter CDF files

NAME	DESCRIPTION	TYPE	COMMENT
Data_product	Type of data product. It shall use the ISTP format "PREFIX>Suffix" (e.g., "HIST1D>1D histogram"), See comment for another example.	P	Note that the descriptor field in the file name is in fact the combination of the instrument [ISTP descriptor attribute] and the Solar Orbiter data_product attribute. Hence, in the example given above for EPD-EPT-ASUN- BURST-ELE-CLOSE, the data_product would be BURST-ELE-CLOSE>Burst Electrons Close Mode
Free_field	Description of the free field in the filename. It shall use the ISTP format "PREFIX>Suffix" (e.g., "NORM>Normal mode"), where the value of the prefix shall correspond to the "free field" in the file naming convention.	P	Only mandatory if the "free field" is provided in the file name.
SOOP_NAME	Name of the SOOP campaign that the data belongs to. E.g., 'L_FULL_HRES_MCAD_Coronal_He_Abundance' See [SAP] and https://issues.cosmos.esa.int/solarorbiterwiki/display/SOSP/SOOP+pages	P	It shall be one entry per SOOP, all L1, L2 and L3 files should be associated with at least one SOOP_NAME; if not, put "none". See notes for SOOPNAME in FITS metadata in Section 3.1.1.4
SOOP_TYPE	SOOP campaign identifier, following the SOOP naming convention described in the [IOR ICD]	P	It shall be one entry per SOOP, all L1, L2 and L3 files must be associated with at least one SOOP_TYPE, otherwise put

NAME	DESCRIPTION	TYPE	COMMENT
			“none”. See notes for SOOPTYPE in FITS metadata in Section 3.1.1.4
OBS_ID	A unique identifier for the observation that is associated with the data acquisition and includes identifying information about: the SOOP; planning period; instrument; and instrument’s observation mode. The format of OBS_ID is defined in [IOR-ICD], and shall be included in the IORs delivered by the Instrument Teams to SOC.	P	If a file is associated with multiple observations, there should be one entry for each OBS_ID. Mandatory for L1, L2 and L3 files. If no OBS_ID then “none”.
Dataset_ID	Data set identifier	O	It can refer to a unique instrument data set identifier.
Process_ID	Process identifier	O	It can refer to a unique process identifier in a instrument data processing pipeline.
Job_ID	Process job identifier	O	It can refer to a unique processed job identifier in a instrument data processing pipeline.
LEVEL	Data processing level as defined in the Solar Orbiter conventions. It shall use the ISTP format "PREFIX>Suffix" (e.g., "L1>Level 1 Data"), where the value of the prefix shall correspond to the "level" field in the file naming convention.	P	Equal to the Data_type keyword value.

3.2.3 *Variable attributes*

Variable attributes are linked with each individual variable and provide additional information about each variable. Solar Orbiter CDF data files shall use the variable attributes listed in the following tables.

3.2.3.1 ISTP Variable Attributes

Table 3-20 ISTP Variable scope attributes for Solar Orbiter CDF Files

NAME	DESCRIPTION	TYPE	CDF DATA TYPE	COMMENT
FIELDNAM	holds a character string (up to 30 characters) which describes the variable.	M	CDF_CHAR	
CATDESC	is an approximately 80-character string which is a textual description	M	CDF_CHAR	

NAME	DESCRIPTION	TYPE	CDF DATA TYPE	COMMENT
	of the variable and includes a description of what the variable depends on.			
DEPEND_0	explicitly ties a data variable to the time variable on which it depends. All variables which change with time must have a DEPEND_0 attribute defined. The value of DEPEND_0 is *EPOCH*, the time ordering parameter for ISTP/IACG.	M	CDF_CHAR	
DEPEND_1	ties a dimensional data variable to a support_data variable on which the 2-th dimension of the data variable depends.	M	CDF_CHAR	Only mandatory for 2d variables
DEPEND_2	ties a dimensional data variable to a support_data variable on which the 3-th dimension of the data variable depends.	M	CDF_CHAR	Only mandatory for 3d variables
DEPEND_3	ties a dimensional data variable to a support_data variable on which the 4-th dimension of the data variable depends.	M	CDF_CHAR	Only mandatory for 4d variables
DISPLAY_TYPE	tells automated software what type of plot to make and what associated variables in the CDF are required in order to do so.	M	CDF_CHAR	Only mandatory for data variable
FILLVAL	is the number inserted in the CDF in place of data values that are known to be bad or missing. Fill data are always non-valid data.	M	Variable data type	Only mandatory for time varying variables. See http://spdf.gsfc.nasa.gov/istp_guide/vattributes.html#FILLVAL for allowed values.
FORMAT	is the output format used when extracting data values out to a file or screen.	M	CDF_CHAR	Only mandatory if not using FORM_PTR

NAME	DESCRIPTION	TYPE	CDF DATA TYPE	COMMENT
LABLAXIS	should be a short character string (approximately 10 characters, but preferably 6 characters - more only if absolutely required for clarity) which can be used to label a y-axis for a plot or to provide a heading for a data listing.	M	CDF_CHAR	Only mandatory if not using LABL_PTR
LABL_PTR_1	is used to label a dimensional variable when one value of LABLAXIS is not sufficient to describe the variable or to label all the axes.	M	CDF_CHAR	Only mandatory if not using LABLAXIS
LABL_PTR_2	is used to label a dimensional variable when one value of LABLAXIS is not sufficient to describe the variable or to label all the axes.	M	CDF_CHAR	Only mandatory if not using LABLAXIS
LABL_PTR_3	is used to label a dimensional variable when one value of LABLAXIS is not sufficient to describe the variable or to label all the axes.	M	CDF_CHAR	Only mandatory if not using LABLAXIS
UNITS	is a character string (no more than 20 characters, but preferably 6 characters) representing the units of the variable.	M	CDF_CHAR	Only mandatory if not using UNIT_PTR. Use the symbols not full words, e.g., T not Tesla, or m/s not metres/second. For variables with no units (e.g., a ratio or a direction cosine) use ‘unitless’.
UNIT_PTR	has as its value a variable which stores the character strings (up to 20 characters per character string) representing the units of the original variable, which can be added to a data listing heading or plot label.	M	CDF_CHAR	Only mandatory if not using UNITS
VALIDMIN	holds value which is the minimum value for a particular variable that	M	Variable data type	Only mandatory for time varying data and support_data

NAME	DESCRIPTION	TYPE	CDF DATA TYPE	COMMENT
	are expected over the lifetime of the mission.			
VALIDMAX	hold values which is the maximum value for a particular variable that are expected over the lifetime of the mission.	M	Variable data type	Only mandatory for time varying data and support_data
SCALEMIN	is value which can be based on the actual value of data found in the CDF data set or on the probable uses of the data, {\em e.g.}, plotting multiple files at the same scale.	M	Variable data type	Only mandatory for time varying data and support_data
SCALEMAX	is value which can be based on the actual values of data found in the CDF data set or on the probable uses of the data, {\em e.g.}, plotting multiple files at the same scale.	M	Variable data type	Only mandatory for time varying data and support_data
VAR_TYPE	identifies a variable.	M	CDF_CHAR	e.g., ‘data’ or ‘support_data’
SCALETYP	indicates whether the variable should have a **linear** or a **log** scale as a default.	M	CDF_CHAR	Only mandatory for data and support_data if not using SCAL_PTR
SCAL_PTR	is used for dimensional variables when one value of SCALTYP is not sufficient.	M	CDF_CHAR	Only mandatory for data and support_data if not using SCALETYP
VAR_NOTES	holds ancilliary information about the variable and can be any length.	M	CDF_CHAR	Can be ‘none’ if no extra details can be added, e.g., for EPOCH.
MONOTON	indicates whether the variable is monotonically increasing or monotonically decreasing.	O	CDF_CHAR	
TIME_BASE	Reference frame where the time is measured (e.g., J2000)	P	CDF_CHAR	Only mandatory for time variables where the data type is not CDF_TIME_TT2000. In that case, it is implicit in the data type
TIME_SCALE	Scale of the time (e.g., Terrestrial Time, UTC)	P	CDF_CHAR	Only mandatory for time variables where the data type is not CDF_TIME_TT2000. In that case, it is implicit in the data type

NAME	DESCRIPTION	TYPE	CDF DATA TYPE	COMMENT
REFERENCE_POSITION	Position in the reference frame where the time is measured (e.g., Rotating Earth Geoid, Spacecraft barycentre, Sun barycentre)	P	CDF_CHAR	Only mandatory for time variables where the data type is not CDF_TIME_TT2000. In that case, it is implicit in the data type
ABSOLUTE_ERROR	Value of the absolute error	O	Variable data type	
RELATIVE_ERROR	Value of the relative error	O	Variable data type	
FORM_PTR	has as its value a variable which stores the character strings (up to 20 characters per character string) representing the desired output format for the original variable.	M	CDF_CHAR	Only mandatory if not using FORMAT

3.2.3.2 Additional Variable Attributes

These additional attributes are not part of the ISTP standard but should be included in Solar Orbiter data files.

Table 3-21 Additional Variable Scope Attributes for Solar Orbiter CDFs

NAME	DESCRIPTION	TYPE	CDF DATA TYPE	COMMENT
SI_CONVERSION	A string that defines the conversion needed to base SI units, e.g., "1.0E-9>T" for DC Magnetic field data in nT.	P	CDF_CHAR	Use the symbols not full words, e.g., T not Tesla, or m/s not metres/second. For variables already given in SI units 1.0>[SI unit], e.g., 1.0>s. For unitless variables, 1>unitless
COORDINATE_SYSTEM	For non-scalar data, contains coordinate system name, e.g., ‘HCI’ or ‘RTN’. Note representation is no longer included.	P	CDF_CHAR	Only mandatory for vector or tensor variables
FRAME_ORIGIN	gives the origin of the reference frame where this is not implicit in	P	CDF_CHAR	Only mandatory for vector or tensor variables

NAME	DESCRIPTION	TYPE	CDF DATA TYPE	COMMENT
	the value of COORDINATE_SYSTEM			
FRAME_VELOCITY	can take either the value 'Observatory' where no corrections have been applied to the data or 'Inertial' where quantities (e.g. electric field or plasma flow velocity) have been corrected for spacecraft motion relative to an inertial frame (HCI)	O	CDF_CHAR	
REPRESENTATION_i	Pointer to a support variable that gives the representation (['x','y','z'] for Cartesian; ['r','p','t'] for spherical polar; ['r','p','z'] for cylindrical polar) of the ith dimension of the variable.	P	CDF_CHAR	Only mandatory for vector or tensor variables
TENSOR_ORDER	Contains the rank or order of a tensor, i.e. 1 for a vector, 2 for a 3x3 tensor.	P	CDF_CHAR	Only mandatory for vector or tensor variables
UCD	UCD keywords as defined by the IVOA	O	CDF_CHAR	http://www.ivoa.net/documents/latest/UCD.html

4 APPENDIX A: VESPA KEYWORDS FOR THE UNIFIED SOLAR ORBITER CATALOGUE

We present here the VESPA (Virtual European Solar and Planetary Access) keywords to be studied as keyword names for the Unified Solar Orbiter Catalogue (USOC). We also propose to link them to existing FITS and CDF metadata, when they already exist.

The units and allowed values proposed in this series of keywords have been selected for a client-server (computer-to-computer) interaction. The actual user interface shall transform the units and values into human readable and usable units and values (e.g.: the spectral range unit for the query is Hz, but the user interface shall allow the user to choose its preferred unit and transform the user query into a protocol query automatically).

The VESPA keyword list proposed here is the released version available on Nov. 2014.

Table 4-1 VESPA Keywords for Solar Orbiter

VESPA keyword	VESPA keyword unit or list of allowed values	VESPA keyword description	SolO-CDF keyword	SolO-FITS keyword	Notes
dataproduct_type	• im • sp • ds • sc • pr • vo • mo • cu • ts • ca • sv	high level scientific organization of the data	DATAPRODUCT_TYPE		See Note [1]
target_name	Sun		TARGET_NAME		
target_class	• star • interplanetary_ medium		TARGET_CLASS		Remote data could have "star" and in-situ "interplanetary_medium", but this may not be adequate.
target_region	• Chromosphere • Corona • Interior • Photosphere • Solar Wind • ...		TARGET_REGION	OBJECT	*The current definition of the FITS "OBJECT" keyword does not fit exactly with this definition. To be mixed with the "Element_name" attribute.* This parameter optionally identifies the region of interest for the resource, in complement to target_name. This parameter only introduces generic regions, not specific local names, which must be handled using the element_name parameter (see examples above). The best practice is to take the values from standard sources:

VESPA keyword	VESPA keyword unit or list of allowed values	VESPA keyword description	SolO-CDF keyword	SolO-FITS keyword	Notes
					• IVOA thesaurus: http://www.ivoa.net/rdf/Vocabularies/vocabularies-20091007/IVOAT/ • IAU thesaurus http://www.mso.anu.edu.au/library/thesaurus/ + another version: http://www.vocabularyserver.com/trex/en/ The latter seems more recent and more complete (although the interface is not practical) • Spase dictionary http://www.spase-group.org/
time_min	Julian day	The date and time of the beginning of the first acquisition for the data contained in the file	TIME_MIN	DATE-BEG	CDF "START_TIME" and FITS "DATE-BEG" are usually in ISO-8601 format, and thus are to be transformed into Julian Day at the time of the USOC table generation
time_max	Julian day	The date and time of the end of the last acquisition for the data contained in the file	TIME_MAX	DATE-END	
time_origin	• target • spacecraft				This a service level, not data product level metadata
time_scale	UTC				This a service level, not data product level metadata
time_sampling_step_min	seconds	Minimum value of the sampling temporal step within the file.			
time_sampling_step_max	seconds	Maximum value of the sampling temporal step within the file.			
time_exp_min	seconds	Minimum value of the exposure time		XPOSURE	

VESPA keyword	VESPA keyword unit or list of allowed values	VESPA keyword description	SolO-CDF keyword	SolO-FITS keyword	Notes
		(=integration time) within the file			
time_exp_max	seconds	Maximum value of the exposure time (=integration time) within the file		XPOSURE	
spectral_range_min	Hz	Minimum value of the spectral range within the file		WAVEMIN	Standard transformation rules from energy, wavelength or wavenumber to frequency are provided in the documentation
spectral_range_max	Hz	Maximum value of the spectral range within the file		WAVEMAX	
spectral_sampling_step_min	Hz	Minimum value of the spectral sampling within the file			This is the spectral separation between the centres of two adjacent filters, channels or pixels.
spectral_sampling_step_max	Hz	Maximum value of the spectral sampling within the file			
spectral_resolution_min	Hz	Minimum value of the spectral resolution within the file			This is the spectral bandwidth used for the measurement (Full Width at Half Maximum). In case of a filter camera this is the filter bandwidth; in case of a spectrometer this is the spectral resolution per se.
spectral_resolution_max	Hz	Maximum value of the spectral resolution within the file			
spatial_frame_type	• celestial • body • cartesian • spherical • healpix	The spatial coordinate flavour used for defining C1, C2 and C3 axes.			See Note [2]

VESPA keyword	VESPA keyword unit or list of allowed values	VESPA keyword description	SolO-CDF keyword	SolO-FITS keyword	Notes
spatial_coordinate_description	String	Name and/or description of the coordinate system (e.g.: HEEC, Carrington, RTN...)		WCSNAME	A list of abbreviations for coordinate systems is being worked on within VESPA project, suggestions very welcome.
spatial_origin	String	Spacecraft or target			
c1min	depends on spatial_frame_type: • m for distances • deg for angles	Minimum value on axis C1 within the file			
c2min		Maximum value on axis C1 within the file			
c3min		Minimum value on axis C2 within the file			
c1max		Maximum value on axis C2 within the file			
c2max		Minimum value on axis C3 within the file			
c3max		Maximum value on axis C3 within the file			
c1_resol_min		Minimum value of the spatial resolution on axis C1 within the file		CDELT1	
c2_resol_min		Minimum value of the spatial resolution on axis C2 within the file		CDELT2	
c3_resol_min		Minimum value of the spatial resolution on axis C3 within the file		CDELT3	
c1_resol_max		Maximum value of the spatial resolution on axis C1 within the file		CDELT1	

VESPA keyword	VESPA keyword unit or list of allowed values	VESPA keyword description	SolO-CDF keyword	SolO-FITS keyword	Notes
c2_resol_max		Maximum value of the spatial resolution on axis C2 within the file		CDELT2	
c3_resol_max		Maximum value of the spatial resolution on axis C3 within the file		CDELT3	
instrument_host_name	Solar Orbiter	Name of the facility/Observatory/ Mission hosting the instrument		OBSRVTRY	Standard systems outside the Solar Orbiter community will look for a internal metadata to find such information.
instrument_name	• MAG • RPW • EPD • SWA • PHI • EUI • SPICE • STIX • Metis • SoloHI	Name of the instrument	Instrument	INSTRUME	This information is in the file name, but standard systems outside the Solar Orbiter community won't decode the filename, but rather look for a internal metadata to find such information.
measurement_type	UCDs	Description of the observed parameter, expressed using UCDs			See Appendix B (Sect. 5) on UCDs.
access_url	URL	Link to the file (at ESAC)			
access_format	CDF or FITS	Format of the file			
access_estsize	kBytes	Size of the file			
preview_url	URL	Link to a preview of the file (format=png)			

VESPA keyword	VESPA keyword unit or list of allowed values	VESPA keyword description	SolO-CDF keyword	SolO-FITS keyword	Notes
element_name	String			OBJECT	Name of the element observed on the Target (e.g., active region number)
reference	bibcode or DOI or URL		REFERENCE		
particle_spectral_type	• energy • mass • mass-per-charge	Type of particle spectral measurements			This defines the units of the particle_spectral_* attributes
particle_spectral_range_min	• eV for energy • amu for mass • amu/q_e for mass-per-charge	Minimum value of particle spectral range in the file			
particle_spectral_range_max		Maximum value of particle spectral range in the file			
particle_spectral_sampling_step_min		Minimum value of particle spectral sampling step in the file			
particle_spectral_sampling_step_max		Maximum value of particle spectral sampling step in the file			
particle_spectral_resolution_min		Minimum value of particle spectral resolution in the file			
particle_spectral_resolution_max		Maximum value of particle spectral resolution in the file			
processing_level	0 to 5	Data processing level of the data in the file			Using CODMAC processing levels.

Notes:
[1] Detailed description of "data product_type" values:

- **im** = image: scalar field with two spatial axes, or association of several such fields, e.g., images with multiple color planes, from multichannel or filter cameras. Maps of planetary surfaces are considered as images.
- **sp** = spectrum: measurements organized primarily along a spectral axis, e.g., a series of radiance spectra.
- **ds** = dynamic_spectrum: consecutive spectral measurements through time, organized as a time series.
- **sc** = spectral_cube: sets of spectral measurements with 1 or 2D spatial coverage, e.g., imaging spectroscopy. The choice between image and spectral_cube is dictated by the characteristics of the instrument (which dimension is most resolved)
- **pr** = profile: scalar or vectorial measurements along 1 spatial dimension, e.g., atmospheric profiles, atmospheric paths, sub-surface profiles...
- **vo** = volume: other measurements with 3 spatial dimensions, e.g., internal or atmospheric structures.
- **mo** = movie: sets of chronological 2D spatial measurements
- **cu** = cube: multidimensional data with 3 or more axes, e.g., all that is not described by other 3D data types such as spectral cube or volume. This is mostly intended to accommodate unusual data with multiple dimensions.
- **ts** = time_series: measurements organized primarily as a function of time (with exception of dynamical spectra and movies, i.e. usually a scalar quantity). Space-borne dust detector measurements are a typical example of a time series.
- **ca** = catalog: can be a list of events, a catalog of object parameters, a list of features.... e.g., a list of asteroid properties. It can be limited to scalar quantities, and possibly limited to a single element (i.e., one catalog entry). Time_series, Profile, and Catalog are essentially tables of scalar values. In Time_series the primary key is time; in Profile it is altitude or distance; in Catalog, it may be a qualitative parameter (name, ID...).
- **sv** = spatial_vector: list of summit coordinates defining a vector, e.g., vector information from a GIS, spatial footprints...

[2] Detailed description of "spatial_frame_type" value:

- **celestial**: 2D angles on the sky, e. g., right ascension c1 and declination c2 + possibly distance from origin c3 – although this is a special case of spherical frame, the order is different.
- **body**: 2D angles on a rotating body: longitude c1 and latitude c2 + possibly a z c3 coordinate. The best practice is to follow the IAU 2009 planetocentric convention [RD12], in particular eastward longitudes and a north pole located on the north side of the invariant plane of the Solar System for planets and satellites (see [RD12] for small bodies, and Annex A for details). The Z coordinate is by default the distance counted from the center of mass.

The spatial_coordinate_description and spatial_origin attributes allow the data provider to indicate different conventions, e. g., to indicate a planetographic frame, or to use altitude above a reference surface as the third coordinate. It is stressed however that using other frames will make comparisons between datasets more difficult.

- **cartesian**: (x,y,z) as (c1,c2,c3). This includes spatial coordinates given in pixels. cylindrical: (r, theta, z) as (c1,c2,c3); angles are defined in degrees.
- **spherical**: (r, theta, phi) as (c1,c2,c3); angles are defined in degrees as in usual spherical systems (E longitude, zenithal angle/colatitude). If the data are related to the sky, "celestial" coordinates with RA/Dec must be used.
- **healpix**: (H, K) as (c1,c2).

5 APPENDIX B: DISCUSSION OF UNIFIED CONTENT DESCRIPTORS

UCD stands for *Unified Content Descriptor*. It is an IVOA (International Virtual Observatory Alliance) standard. Generic documentation can be found here. The introduction of the UCD standard description is recalled here:

The Unified Content Descriptor (UCD) is a formal vocabulary for astronomical data that is controlled by the International Virtual Observatory Alliance (IVOA). The vocabulary is *restricted* in order to avoid proliferation of terms and synonyms, and *controlled* in order to avoid ambiguities as far as possible. It is intended to be flexible, so that it is understandable to both humans and computers. UCDs describe astronomical quantities, and they are built by combining words from the controlled vocabulary.

A UCD does not define the units nor the name of a quantity, but rather "*what sort of quantity is this?*"; for example **phys.temperature** represents a temperature, without implying a particular unit. It would be possible to describe astronomical data quantities in a natural language such as English or Hungarian or Uzbek; however, it would be very difficult to expect a machine to 'understand' it in any sense. At the opposite extreme, there is an attempt within the IVOA to describe astronomical data in terms of a hierarchical data model, so that there is a place for everything, and everything is in its place. The UCD vocabulary falls between these extremes and is intended to be understandable to both humans and computers. Other explanations can be found at CDS.

There are other standards available for describing the quantity related to a parameter in a data product. For instance, there is the SPASE (Space Physics Archive Search and Extract) "Measurement Type" keyword, which is well adapted for space physics measurement, but still very limited. There is also a series of keywords in the NASA/PDS data dictionary that could be used here, but it is not a stable standard.

5.1 Building UCDs

A UCD is a string of characters. The UCD syntax is described in the UCD IVOA documentation.

There are several online resources to help the user or data provider selecting the correct UCD value.

- Current list of UCD (v1.23)
- CDS UCD builder (based on syntax analysis)
- GAVO UCD resolver (based on existing UCD/description occurrences)

NB When trying to build UCDs, it is very common to try to set up a very precise and accurate description of the parameter. This is not the goal of UCDs. The description should stay at a very simple level. Trying to go into too much detail usually results in trying to use the UCD concept as a data model for your physical parameters. As UCDs were not meant for this, it doesn't work.

5.2 Example of UCDs

Table 5-1 Example UCDs

Instrument	Parameter Name	Parameter Description	UCD	Notes
EUI	EUI-HRI-17nm	EUI-HRI-17nm measures the photon count per pixel in a EUV band around 17nm.	phot.count;em.UV.10-50nm	check on CDS UCD builder with query "photon count at 17 nm"
		At higher data levels we will normalize with respect to the exposure time	phot.flux;em.UV.10-50nm	
RPW	RPW-HFR-L2	Spectral flux density in the radio range (<20 MHz) using electrical sensors	phys.flux.density;phys.electField;em.radio	