

Provenance Metadata

On this page...

- [Data Provenance](#)
 - [Header Keywords](#)
 - [FITS Provenance Extension](#)
 - [Table Fields](#)
- [For Further Reading...](#)

Data Provenance

Establishing the provenance of science data products is important to the users of HLSP collections for a number of reasons, including:

- Promoting data traceability and reproducibility
- Establishing the pedigree for the quality of the data processing
- Understanding the facilities, instruments, configurations used obtain the data, and under what environmental circumstances the data were obtained

Provenance should be recorded in multiple ways, including the Project Description file and the README file that must be included in every HLSP collection (see [Required Contents](#)). It is also established in the primary journal paper where each HLSP collection is described. Various of the metadata listed in this chapter as required or recommended partially address these needs. The focus of this article is the recording of sufficient metadata that associations can be created between HLSP products and the MAST mission products (or the products from other observatories) from which they were derived. Two approaches have been used for this purpose, as described in the subsections below.

Header Keywords

The other sections of this chapter identified some keywords where the value in the contributing products may be different. To represent a meaningful value in the HLSP product:

- the value should be 'MULTI'
- the keyword record should be followed by supplemental records with an abbreviation of the original keyword, with a 2-digit numerical suffix
 - each supplemental record should contain a value appropriate to the contributing product, in numerical order

The following table illustrates the concept for a composite UV spectrum of a target, using two HST instruments. Note that the contributing observations are linked to the HLSP product through the MAST Observation ID, shown in the table as 'FILEIDnn'.

Keyword	Value
DISPERSR	'MULTI'
DISPER01	'G130M/180'
DISPER02	'E230M'
FILEID01	'1ea22cx3q'
FILEID02	'ockp07020'
INSTRUME	'MULTI'
INSTRU01	'COS'
INSTRU02	'STIS'



An acceptable synonym for FILEIDnn is DATAⁿⁿ.

FITS Provenance Extension

It is sometimes tedious, even impractical, to record critical metadata for every data product that contributed to constructing an HLSP product. Examples include:

- combined images constructed from many exposures,
- multi-band catalogs that draw data from many observing facilities,
- SEDs composed from spectra from multiple observing facilities, telescopes, and instruments in one or more configurations

In these cases it may be better to collect this information in a BINTABLE extension to each HLSP FITS product. The table should have the following attributes:

- Specify EXTNAME = 'PROVENANCE' in the extension header
- List each contributing data product, one per table row
- Use one table column per attribute
- At a minimum, include all relevant attributes that vary among contributing products
- Specify the data type and units (if applicable) in the header for each attribute

Table Fields

The following table gives examples of attributes that may be applicable to an HLSP collection.

Field Name	Example Value	Description
FILE_ID	'ockp11020'	File name or observatory-unique identifier of the contributing observation. For products from MAST missions, provide the Observation ID so that the contributing data may be linked within MAST.
DATE-BEG	'2021-01-05T12:34:56.78'	ISO 8601-formatted date-time string for observation start
DATE-END	'2021-01-05T12:39:56.78'	ISO 8601-formatted date-time string for observation end
DISPRSR	'G140L'	Name of dispersing optical element used
FILTER	'F25ND'	Name of (possibly passband limiting) filter used
INSTRUMENT	'STIS'	Name of instrument used
RADESYS	'ICRS'	Coordinate reference frame
TELESCOPE	'HST'	Name of telescope used
XPOSURE	300	Total duration of exposure in sec, exclusive of dead-time

For Further Reading...

- [Common Metadata](#)