CERES Software Bulletin 97-06

Revision 2 - January 27, 1998

CERES Metadata Approach

Maria Vallas Mitchum (m.v.mitchum@larc.nasa.gov)
Alice Fan (t.f.fan@larc.nasa.gov)

1.0 Purpose

This bulletin describes the CERES metadata policy, approach, and functionality.

2.0 Background

The Earth Observation System Data Information System Core System (ECS) data model objects are described through the Earth Science Data Type (ESDT) descriptor file for each data collection (product). The Metadata Configuration File (MCF) is a component, or subset, of the ESDT descriptor file and it specifies the set of metadata attributes chosen to describe a granule (an instance of a product). The fields of the MCF are populated during product generation by the Product Generation Executive (PGE) software. The populated fields of the MCF are written both to a Hierarchical Data Format (HDF) data product and a metadata load file (.met), which is subsequently inserted into Science Data Server to populate the inventory database tables.

3.0 CERES Metadata Policy

The ECS metadata tools do not automatically write metadata to a non-HDF file. The CERES Data Management Team mandated that a set of selected attributes be written, in ASCII format, to the header of all Direct Access files. It was also decided that a selected subset of the attributes, in Vdata format, be written to HDF and HDF-EOS files. Justification is that both sets of metadata provide information to users without the necessity of the PDPS Toolkit library.

The CERES metadata policy is as follows:

1) A .met file, in Object Description Language (ODL) format, will be generated for all non-temporary output files.
2) The complete set of CERES metadata, in ODL format, and a selected subset of CERES metadata, in Vdata format, will be written to HDF and HDF-EOS data files.
3) The complete set of CERES metadata, in ASCII format, will be written to the header of Direct Access files.
4) Optional: User could write a second Vdata to an HDF or HDF-EOS data file and a second header to a Direct Access data file.
5) NO header metadata will be written on ASCII or binary Sequential Access output files, although their respective .met files will be generated.

4.0 CERES Baseline Header Metadata

There are 41 attributes identified as the CERES Baseline Header Metadata. Appendix A lists these attributes with their source and data type. Please refer to the computer bulletin 97-12, "CERES Metadata Requirements for LaTIS", for a detail description of these attributes. They have been categorized into two groups: Inventory, which are the searchable items that will be written on the file header and will populate the DataServer Database, and Archive, which will be stored on the header, exclusively, to provide additional information while reading the data file.

5.0 Metadata Wrapper Approach

A Fortran 90 metadata wrapper has been written, which calls Toolkit functions, to generate the CERES metadata. This wrapper was designed to shield the CERES team members from dealing with the evolving data model, metadata tools, and MCF structure and requirements.

There are several input sources from which the metadata wrapper will extract the required metadata. The sources are: (1) the Process Control File, PCF, (2) the software generated metadata, (3) a select set of system parameters, and (4) MCF.

5.1 Process Control File

PCF metadata parameters will be stored in the Runtime Parameter Section. Several will be chosen at the PGE Production Request time. These are: CERPGEName, SamplingStrategy and ProductionStrategy (see Appendix B), Data Year, Data Month, Data Day, and Hour of Month. These parameters and other PGE dependent Runtime Parameters are input to the production system and saved in the database from the Production Request submitted by the Science Team. They are retrieved from the database and inserted into the PCF at the PCF generation time.

The Configuration Code, Software System Configuration Change Request (SCCR), and Ancillary Data SCCR numbers will be retrieved by the Production System from the Configuration Management (CM) system database, based on the name of the PGE, CERPGEName, and stored on the PCF at the PCF generation time.

5.2 Software generated metadata

The Subsystem software is required to provide a minimum of 10 attributes:

1. CERES Bounding rectangle or Gring in 0-180 and 0-360 ranges (4 attributes)
2. BeginningDate, BeginningTime, EndingDate, and EndingTime (4 attributes)
3. AutomaticQualityFlag and AutomaticQualityFlagExplanation (2 attributes)

5.3 System parameters

The system parameters accessed by the wrapper include the computer’s name and operating system, system clock, and version numbers of the Toolkit, HDF, HDF-EOS, and CERESlib.

5.4 Metadata Configuration File

Each output product will have a unique MCF, in ODL format. The MCF is a template containing the product unique metadata field attributes, predefined for the product by the

Subsystem Working Group (SWG). This file is used as input by the wrapper to accomplish the Toolkit metadata functions. The ECS ShortName, in the MCF, is used to identify the data collection for the output product granule. The MCF will be supplied by the Metadata Team (Ms. Maria Mitchum and Ms. Alice Fan) for each output product.

Additionally, there is a C version of the metadata wrapper tailored to the needs of the Instrument Subsystem (SS 1.0) which interfaces with Ada software programs.

6.0 Metadata Wrapper Methodology

The method in which metadata is written on an HDF/HDF-EOS file and Direct Access file differ.

6.1 HDF/HDF-EOS

All of the CERES baseline metadata (see Appendix A) and Product Specific Attributes (PSAs), see Section 8.0, will be written to HDF and HDF-EOS files using Toolkit calls in the form required to interface to the ECS and Langley TRMM Information System (LaTIS) systems. In addition to this, a Vdata table of 14 (out of the baseline 41) selected attributes will be written to the file for non-Toolkit users. If additional information is required by the SWG, a second Vdata may be written by the Subsystem software.

6.2 DIRECT ACCESS BINARY:

All of the CERES baseline metadata and PSAs will be written to Direct Access files as an ASCII header record. (Note: No header records will be written on Sequential Access binary files or ASCII files.) Each attribute takes 80 bytes. Users are required to reserve enough bytes at the beginning of the data file or the wrapper will overwrite the internal data. This means that one or more records must be reserved for the wrapper at the beginning of each file.

Like the HDF files, developers might want to write their own header information as a second header. The following formula should be used to calculate the number of bytes required for the header record, which should be compared with your individual product's record size. Please be aware that this, at best, is a rough estimate.

(41 attributes + # of InputPointers * records per InputPointer + # of PSA) * 80 bytes

Please see Appendix C for an example partial listing of metadata in an HDF file and Appendix D for an example listing of metadata in a Direct Access file.

7.0 Metadata Wrapper Functionality

As stated above, one of the functions of the wrapper is to produce a unique metadata file for the DataServer Database containing information for each output file. The metadata file is created following the MCF format, written in ODL, using the identical filename with a .met extension.

Internally the wrapper has the following functions:

1. Convert the CERES coordinates into ECS valid range values.
2. Populate the value for the InputPointer attribute by collecting OPENed input file pointers from the CERES Software Library, CERESlib, I/O module.
3. Derive values of instrument and platform shortnames from the SamplingStrategy attribute.
4. Construct LocalVersionID attribute to include the information about the software library versions used within the software.
5. Construct ProductGenerationLOC attribute to include the information about the computer and operating system.
6. Retrieve the entire set of attributes, or a particular attribute, for a specified input product, which has a .met file, or header metadata.

The details of how to use the wrapper to write metadata are addressed in computer bulletin 97-08: "How to Write Metadata to a Granule Using the Wrapper". The details of how to use the wrapper to read metadata are addressed in computer bulletin 97-10: "How to Read Metadata from a Granule Using the Wrapper".

8.0 Additional Options for Each Product

Each Software program can provide more product specific attributes through the Product_Specific_Attribute array, which will be tailored for each output product. If the PSAs are to be stored in the Inventory group, then each will fall in one of two categories:

1. If PSA has been established as a core attribute by ECS: good to go.
2. If PSA is new: will need to be preprocessed to establish its existence and definition. This must be brought to the attention of the Metadata Working Group.

If the PSA is to be stored with the Archive metadata, no previous definition to ECS is required.

Appendix A: CERES Baseline Header Metadata

The following table describes the metadata that will be written on the CERES Header record for all output products. In order to read the table the following notations have been used:

* optional attributes and provided by subsystem software.
** required attributes and provided by subsystem software.
v Vdata attributes

where: s = string, F = float, I = integer, datetime = yyyy-mm-ddThh:mm:ss.xxxxxxZ
Inv = Inventory Metadata, Arc = Archive Metadata

item	Source	Attribute Name	Data Type	Inv/Arc
1. v	ESDT/MCF	ShortName	s(8)	Inv
2.	ESDT/MCF	VersionID	I3	Inv
3.	Delivered to DAAC with PGE	CERPGEName(PSA)	s(20)	Inv.
4.	PCF	SamplingStrategy(PSA)	s(20)	Inv
5.	PCF	ProductionStrategy(PSA)	s(20)	Inv
6.	PCF	CERDataDateYear(PSA)	s(4)	Inv
7.	PCF	CERDataDateMonth(PSA)	s(2)	Inv
8.	PCF	CERDataDateDay(PSA)	s(2)	Inv
9.	PCF	CERHrOfMonth(PSA)	s(3)	Inv
10. v	PGE**	RangeBeginningDate	date	Inv
11. v	PGE**	RangeBeginningTime	time	Inv
12. v	PGE**	RangeEndingDate	date	Inv
13. v	PGE**	RangeEndingTime	time	Inv
14. v	PGE**	AutomaticQualityFlag	s(64)	Inv
15. v	PGE**	AutomaticQualityFlagExplanation	s(255)	Inv
16.	PGE*	QAGranuleFilename(PSA)	s(255)	Inv
17.	PGE*	ValidationFilename(PSA)	s(255)	Inv
18.	wrapper	EastBoundingCoordinate	F11.6	Inv

item	Source	Attribute Name	Data Type	Inv/Arc
19.	wrapper	NorthBoundingCoordinate	F11.6	Inv
20.	wrapper	SouthBoundingCoordinate	F11.6	Inv
21.	wrapper	WestBoundingCoordinate	F11.6	Inv
22.	wrapper	GRingPointLatitude	F11.6	Inv
23.	wrapper	GRingPointLongitude	F11.6	Inv
24.	wrapper	GRingPointSequenceNo	I5	Inv
25.	wrapper	ExclusionGRingFlag	s(1)	Inv
26. v	wrapper	AssociatedPlatformShortName	s(20)	Inv
27. v	wrapper	AssociatedInstrumentShortName	s(20)	Inv
28.	wrapper	ImagerShortName(PSA)	s(20)	Inv
29. v	PCF	LocalGranuleID	s(80)	Inv
30. v	wrapper	LocalVersionID	s(60)	Inv
31.	wrapper	InputPointer	s(255)	Inv
32.	wrapper	PGEVersion	s(10)	Inv
33.	PGE**	CEREastBoundingCoordinate	F11.6	Arc
34.	PGE**	CERNorthBoundingCoordinate	F11.6	Arc
35.	PGE**	CERSouthBoundingCoordinate	F11.6	Arc
36.	PGE**	CERWestBoundingCoordinate	F11.6	Arc
37.	PGE**	CERGRingPointLatitude	F11.6	Arc
38.	PGE**	CERGRingPointLongitude	F11.6	Arc
39. v	wrapper	CERProductionDateTime	datetime	Arc
40. v	PGE*	NumberofRecords	I10	Arc
41. v	wrapper	ProductGenerationLOC	s(255)	Arc

Appendix B: Filenaming Convention

The filename of each output file will be stored as an attribute in the metadata parameter LocalGranuleName, a product specific attribute. The LocalGranuleName will follow Dr. Bruce Barkstrom's Naming Convention. Reference the document, "CERES Metadata Requirements for LaTIS", for a detail description. Data Type: (s80)

[Investigation]_ [Product-ID]_[SamplingStrategy]_[ProductionStrategy]_[Configuration].[Instance]

Note: Field separators will be the underscore (_) except for the last field, where a period (.) is required before the Instance.

Source of LocalGranuleName parameters:

Investigation: CER (fixed)

Product-ID: For Archival Products, the Data Product Catalog (DPC) name will be used, i.e. SSF, else the ESDT Shortname will be modified and used as the product's name. Source: PGE argument, input to the metadata wrapper.

SamplingStrategy: Derived from a LaTIS table at PGE request time, and used as a runtime parameter in the PCF. This is a description of the data source, which typically uses the satellite, instrument combination and imager source that contributes to the product. Valid Values: {TRMM-PFM-VIRS, EOS-AM1-MODIS} Data Type: (s20)

ProductionStrategy: Derived from a LaTIS table at PGE request time, and used as a runtime parameter in the PCF. Valid Values: {PreFlight, Edition, Campaign, DiagnosticCase, TBD} Data Type: (s20)

Configuration: Latest configuration code #. Source: PCF runtime parameter for the subsystem, will be retrieved from the LaTIS database table.

Instance: A variable length identifier chosen by the working group to uniquely identify the instance in the sampling strategy. If the identifier includes a data date, it must be of the form YYYYMM[DD][HH], such as 1997111501 or 20000312. Less commonly, the Instance may include spatial identifiers, such as Zone numbers or latitude bands. Thus, we might have 199903zone180. In most cases this parameter will resort to a default value. Source: PGE generated.

Example: CER_BDS_TRMM-PFM_PreFlight_00001.19970729

Appendix C: HDF Metadata in ODL Format (Partial Listing)

```
GROUP                  = INVENTORYMETADATA
  GROUPTYPE            = MASTERGROUP
  GROUP                  = ECSDATAGRANULE
    OBJECT                 = PRODUCTIONDATETIME
      NUM_VAL              = 1
      VALUE                = "NOT OBTAINED"                 /* set by TK */
    END_OBJECT             = PRODUCTIONDATETIME
  END_GROUP              = ECSDATAGRANULE

  GROUP                  = MEASUREDPARAMETER
    OBJECT                 = MEASUREDPARAMETERCONTAINER
      CLASS                = "1"
      GROUP                  = QAFLAGS
        CLASS                = "M"
        OBJECT                 = AUTOMATICQUALITYFLAG
          NUM_VAL              = 1
          CLASS                = "1"
          VALUE                = "Passed"
        END_OBJECT             = AUTOMATICQUALITYFLAG
        OBJECT                 = AUTOMATICQUALITYFLAGEXPLANATION
          NUM_VAL              = 1
          CLASS                = "1"
          VALUE                = "no error detected"
        END_OBJECT             = AUTOMATICQUALITYFLAGEXPLANATION
      END_GROUP              = QAFLAGS
    END_OBJECT             = MEASUREDPARAMETERCONTAINER
  END_GROUP              = MEASUREDPARAMETER

  GROUP                  = COLLECTIONDESCRIPTIONCLASS
    OBJECT                 = SHORTNAME
      NUM_VAL              = 1
      VALUE                = "CGSSF_AB"
    END_OBJECT             = SHORTNAME
    OBJECT                 = VERSIONID
      NUM_VAL              = 1
      VALUE                = 1
    END_OBJECT             = VERSIONID
  END_GROUP              = COLLECTIONDESCRIPTIONCLASS
  GROUP                  = INPUTGRANULE
    OBJECT                 = INPUTPOINTER
      NUM_VAL              = 800
      VALUE                = ("./96097210014i09.B1D", "./96097210014i09.abc")
    END_OBJECT             = INPUTPOINTER
  END_GROUP              = INPUTGRANULE


Note: This is a partial listing, Page 1 of 7.
```

Appendix D: ASCII header with Bounding Rectangle Example

```
BEGIN_HEADER
ShortName                              = CGFLATAB
VersionID                              =     1
CERPGEName                             = 4.6P1
SamplingStrategy                       = TRMM-PFM-VIRS
ProductionStrategy                     =  AtLaunch
CERDataDateYear                        = 1996
CERDataDateMonth                       = 01
CERDataDateDay                         = 15
CERHrOfMonth                           = 352
RangeBeginningDate                     = 1996-01-15
RangeBeginningTime                     = 15:00:00.000000
RangeEndingDate                        = 1996-01-15
RangeEndingTime                        = 16:00:00.000000
AssociatedPlatformShortName.1          = TRMM
AssociatedInstrumentShortName.1        = PFM
LocalGranuleID                         =  CER_SSFB_TRMM-PFM-VIRS_AtLaunch_00001.1996011515
PGEVersion                             = 00001
CERProductionDateTime                  = 1998-01-08T12:32:21.000000Z
LocalVersionID                         =  Clib-19971230  TK5.2.1 HDF-4.1r1  HDFEOS- 2.0  SW00013  ANC00015
ProductGenerationLOC                    = NASA Langley Research Center, HOST - thunder1-f  OS -IRIX64
NumberOfRecords                        =      9876
WestBoundingCoordinate                 =   60.000000
NorthBoundingCoordinate                =   90.000000
EastBoundingCoordinate                 = -140.000000
SouthBoundingCoordinate                =  -90.000000
CERWestBoundingCoordinate              =   60.000000
CERNorthBoundingCoordinate             =    0.000000
CEREastBoundingCoordinate              =  220.000000
CERSouthBoundingCoordinate             =  180.000000
AutomaticQualityFlag.1                 = Passed
AutomaticQualityFlagExplanation.1      = no error detected
ImagerShortName                        = VIRS
CERHrOfDay                             =        15
InputPointer.1   =  /disk2/thunder/fan/Meta1/CER_SSFI_TRMM-PFM-VIRS_AtLaunch_00001.1996011515
InputPointer.2   =  /disk2/thunder/fan/Meta1/CER_LWSM_TRMM-PFM_ArLaunch_00014.1996Winter
END_HEADER
```