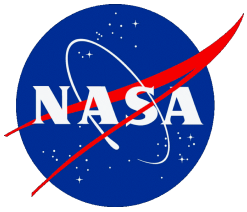
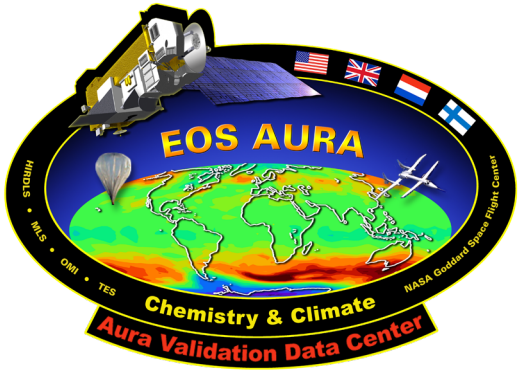


National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, MD



Addendum to the “Generic metadata guidelines on atmospheric and oceanographic datasets for the Envisat Calibration and Validation Project” as implemented by the Aura Validation Data Center (AVDC)

NASA GEST Work Activity 610-75-262

NASA NIWA-ERI Contract# NAS05903

Original: August 8, 2006
Revised: August 31, 2006
Version: August 31, 2006

Document Profile Information

AVDC_file_format_addendum_20060831.doc	
Title:	Addendum to the "Generic metadata guidelines on atmospheric and oceanographic datasets for the Envisat Calibration and Validation Project" as implemented by the Aura Validation Data Center (AVDC)
Version:	August 31, 2006
Type:	Guidelines/Data reporting requirements
Audience:	Data providers
Author(s):	Bojan R. Bojkov (GEST and NASA-GSFC, bojan.bojkov@gsfc.nasa.gov), Ian Boyd (NIWA, i.boyd@niwa.com), Martine De Mazière (BIRA-IASB, Martine.DeMaziere@bira-iasb.oma.be), and Robert M. Koopman (ESA-ESRIN, Rob.Koopman@esa.int)
Status:	Final
Distribution:	Public
Location:	
Abstract:	This document describes the modifications to and the implementation of the European Space Agency's "Generic metadata guidelines on atmospheric and oceanographic datasets for the Envisat Calibration and Validation Project" document by the Aura Validation Data Center (AVDC).
See Also:	"Generic metadata guidelines on atmospheric and oceanographic datasets for the Envisat Calibration and Validation Project" by Bojkov *et al.* (2002)

Table of Contents

1 Overview

This addendum describes modifications[1] to the European Space Agency's "Generic metadata guidelines on atmospheric and oceanographic datasets for the Envisat Calibration and Validation Project" document (Bojkov *et al.*, 2002). Specifically, updates to the global and variable metadata attributes, as well as the implementation of the Hierarchical Data Format (HDF4 and HDF5), by the NASA Aura Validation Data Center (AVDC) are outlined.

2 Metadata attribute modifications

The Aura Validation Data Center (AVDC) has implemented four metadata attribute modifications into the original Envisat metadata guidelines (Bojkov et al., 2002). These metadata attributes changes are summarized in Table 2, while a more detailed description of each attribute is given in this chapter's sub-sections.

Table 2: Summary of metadata changes.

Attribute Name	Attribute Type	Change	Comment
DATA_TYPE	Global Attribute	DATA_LEVEL	*New name for clarity*
DATA_FILE_VERSION	Global Attribute	Additional entry formats allowed	*Now can also describe processing version. For example v8, 5.01, etc.*
FILE_META_VERSION	Global Attribute	Requires 2 mandatory entries	*Defined metadata versioning and tool names attribute entries*
VAR_MONOTONE	Variable Attribute	Removed	
VIS_SCALE_TYPE	Variable Attribute	Entry change	*If VIS_PLOT_TYPE set to NONE, then VIS_SCALE_TYPE must be set to NONE;NONE*
VIS_SCALE_MIN/MAX	Variable Attribute	Entry change	*If VIS_PLOT_TYPE set to NONE, then VIS_SCALE_MIN and VIS_SCALE_MAX must be set to NONE*

2.1 DATA_LEVEL (formerly DATA_TYPE)

The global attribute **DATA_LEVEL** (formerly **DATA_TYPE**) specifies the data granularity with respect to time and the data product level. These specifications are concatenated into a single field entry.

Type: STRING, maximum 2 characters
Entry: Single concatenated entry
Format: Time granularity code // Data level code
Example: DATA_LEVEL = H2 … for hourly level 2 data

[1] Due to the rapidly changing nature of the validation metadata, this document does not cover specific attribute field changes and/or additions (i.e. new instrument names, entities, etc). Up-to-date attribute field values are available from the AVDC web site (http://avdc.gsfc.nasa.gov).

Table 2.1a: Allowed "time granularity" codes to construct the **DATA_LEVEL** attribute.

Granularity code	Granularity of the order of…
D	*Days*
H	*Hours*
M	*Minutes*
S	*Seconds*
O	*Other…*

Table 2.1b: Allowed "data level" codes to construct the **DATA_LEVEL** attribute.

Data level code	Comment
0	*Reformatted, time-ordered instrument data*
1	*Geolocated, radiometrically and/or spectrally calibrated instrument data*
2	*Extracted geolocated geophysical data*
3	*Added-value/derived geophysical data, typically gridded data*
4	*Assimilated geophysical data*

2.2 DATA_FILE_VERSION

The global attribute **DATA_FILE_VERSION** specifies the version of the data. Two formulations are permitted for this attribute:

If **DATA_FILE_VERSION** refers to a scientific algorithm or specific data processing version, the entry must follow the format: ddd.ddda, where "d" represents a numerical digit and "a" is one of the 26 letters of the alphabet. For example, for the Envisat SCIAMACHY level 1 data version 4.02b, the attribute entry would become 4.02b; for the reprocessed TOMS version 8 ozone column data, the entry would become 8.

If **DATA_FILE_VERSION** is not associated with a scientific algorithm or a processing algorithm, the attribute entry specifies an arbitrary version of the file, beginning with 001 (with leading zeroes). With each update the data file version should by incremented by 1.

Type: STRING
Entry: Single field
Format: ddd.ddda or ddd including leading zeroes.
Example 1: DATA_FILE_VERSION = 4.02b
Example 2: DATA_FILE_VERSION = 003

2.3 FILE_META_VERSION

The global attribute **FILE_META_VERSION** indicates the version of the metadata definitions used in the data file and the tool name used to generate the current HDF data file.

Type: STRING
Entry: **Two** semicolon-separated fields

Format: 02Rddd; tool name
Example: FILE_METAVERSION = 02R001; ASC2HDF

Table 2.3: Example tool names for the **FILE_META_VERSION** attribute.

Tool name	*Comment*
ASC2HDF	*NILU's ASCII to HDF converter*
CR8HDF	*NASA's ASCII 2 HDF tool suite*
CUSTOM	*Customized tool*

2.4 VAR_MONOTONE

The variable attribute **VAR_MONOTONE** was removed from the required variable attributes in AVDC type HDF files. This variable attribute was ill suited for Earth Science datasets, especially for datasets with variables of dimension two and larger.

2.5 VIS_SCALE_TYPE

If the variable attribute **VIS_PLOT_TYPE** entry is set to **NONE**, then the variable attribute **VIS_SCALE_TYPE** must be set to NONE;NONE.

2.6 VIS_SCALE_MIN/MAX

If the variable attribute **VIS_PLOT_TYPE** entry is set to **NONE**, then the variable attributes **VIS_SCALE_MIN** and **VIS_SCALE_MAX** entries must also be set to NONE.

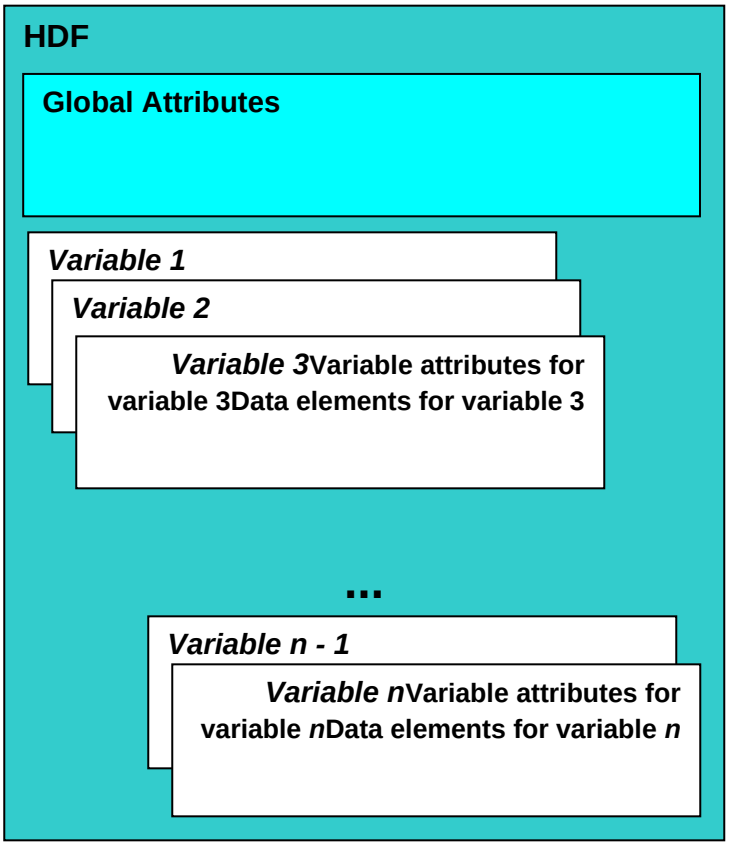
2.7 Other

The variable attribute entry "**FALSE**" for the variable attributes **VAR_AVG_TYPE**, **VIS_PLOT_TYPE**, **VIS_SCALE_TYPE**, **VIS_SCALE_MIN** and **VIS_SCALE_MAX** has been replaced by "NONE" for clarity.

3 HDF Implementations

The AVDC/Envisat HDF implementation consists of data and the associated metadata. The metadata parameters are divided into Global Attributes (pertaining to an entire dataset contained in one single file), and Variable Attributes (pertaining to one single variable within a dataset). A schematic of the AVDC HDF4 (also for HDF5) structure is given in Figure Error: Reference source not found.

Figure Error: Reference source not found: Overview of the AVDC/Envisat HDF structure.



3.1 Global Attributes

To organize the global attributes, they have been grouped into three categories, namely **Originator Attributes**, **Dataset Attributes** and **File Attributes**. Each one of the global attributes appears once in each data file. All attributes marked mandatory in Table 3.1 must have an entry. For the ones labeled optional the entry may remain empty.

Table 3.1: List of mandatory Global Attributes to be included in each file.

Originator Attributes	Required	Comment
PI_NAME	X	
PI_AFFILIATION	X	
PI_ADDRESS	X	
PI_EMAIL	X	
DO_NAME	X	
DO_AFFILIATION	X	
DO_ADDRESS	X	
DO_EMAIL	X	
DS_NAME	X	
DS_AFFILIATION	X	
DS_ADDRESS	X	
DS_EMAIL	X	
Dataset Attributes		
DATA_DESCRIPTION	X	*Free format*
DATA_DISCIPLINE	X	*Refer to standard*
DATA_GROUP	X	*Refer to standard*
DATA_LOCATION	X	*Refer to standard*
DATA_SOURCE	X	*Refer to standard*
DATA_LEVEL	X	*Refer to standard*
DATA_VARIABLES	X	
DATA_START_DATE	X	*ISO8601*
DATA_FILE_VERSION	X	
DATA_MODIFICATIONS	X	*Free format*
DATA_CAVEATS	O	*Free format*
DATA_RULES_OF_USE	O	*Free format*
DATA_ACKNOWLEDGEMENT	O	*Free format*
Originator Attributes		
FILE_NAME	X	*Naming convention*
FILE_GENERATION_DATE	X	*ISO8601*
FILE_ACCESS	X	*Project dependent*
FILE_PROJECT_ID	O	*Project dependent*
FILE_ASSOCIATION	O	*Project dependent*
FILE_META_VERSION	X	*Refer to standard*

3.2 Variable Attributes

Unlike the global attributes, the variable attributes are the specific metadata for a single variable. Each variable listed in the global attribute DATA_VARIABLES must be accompanied by a complete set of associated variable attributes. The variable attributes are grouped into two categories, namely the **Variable Description Attributes** whose names begin with "VAR_", and the **Variable Visualization Attributes** whose names begin with "VIS_". Table 3.2: gives an overview of all variable attributes.

Table 3.2: List of mandatory Variable Attributes that accompany each dataset.

Variable Description Attributes	Required	Comment
VAR_NAME	*X*	*Refer to standard*
VAR_DESCRIPTION	*X*	*Free format*
VAR_NOTES	*O*	*Free format*
VAR_DIMENSION	*X*	
VAR_SIZE	*X*	*The number of elements in each dimension*
VAR_DEPEND	*X*	*INDEPENDENT, CONSTANT or a previously written one dimensional variable*
VAR_DATA_TYPE	*X*	*Allowable formats are INTEGER, LONG, REAL, DOUBLE*
VAR_UNITS	*X*	*Refer to standard for permissible units*
VAR_SI_CONVERSION	*X*	*Refer to standard*
VAR_VALID_MIN	*X*	
VAR_VALID_MAX	*X*	
VAR_AVG_TYPE	*X*	*Refer to standard*
VAR_FILL_VALUE	*X*	*Needs to be outside VAR_VALID_MIN and VAR_VALID_MAX values*
Variable Visualization Attributes		
VIS_LABEL	*X*	*Free format*
VIS_FORMAT	*X*	*Needs to accommodate valid minimum, valid maximum and the fill values*
VIS_PLOT_TYPE	*X*	*Refer to standard*
VIS_SCALE_TYPE	*X*	*Refer to standard*
VIS_SCALE_MIN	*X*	*Refer to standard*
VIS_SCALE_MAX	*X*	*Refer to standard*
HDF4 SDS Pre-defined Attributes		*For HDF4 type files only*
long_name		*From VIS_LABEL*
Units		*From VAR_UNITS*
Format		*From VIS_FORMAT*
Coordsys		*From VIS_SCALE_TYPE*
valid_range		*From VAR_VALID_MIN and VAR_VALID_MAX*
Fillvalue		*From VAR_FILL_VALUE*

3.3 HDF4 Structure

The AVDC and Envisat HDF4 compliant files use the HDF SDS (Scientific Data Structure) data model. The SDS allows for the efficient implementation of the Global Attributes and up to 64 variables of rank 8 with their corresponding Variable Attributes (NCSA, 1999).

Data and Metadata are written to file using the HDF4 SDS interface routines. Each dataset has a unique identifier in the file. Pre-defined variable attributes, listed in Table 3.2, are written using:

- the SDsetdatastrs/setrange/setfillvalue procedures (or the IDL/Fortran equivalents);

- the SDsetdimstrs procedure (or IDL/Fortran equivalents), when the VAR_DEPEND dimension values reference previously written datasets.

3.4 HDF5 Structure

The National Center for Supercomputing Applications (NCSA) HDF5 type file is conceptually related to HDF4, but it uses a completely new format and library (Baker, 2005). HDF5 has a simpler data model that includes two primary objects: datasets and groups. Global Attributes and datasets are written to the root group and the variable attributes[1] are associated with the datasets.

The NCSA provides utilities that can convert HDF4 files to HDF5, and vice versa, but they are generic in nature. The AVDC HDF5 format takes, as a starting point, the resulting output when converting an AVDC HDF4 file to HDF5 using the NCSA utility h4toh5.exe[2]. The VAR_DEPEND attribute information is included by means of an object reference pointer attached as an attribute to the dataset, as shown in Figure 3.4 (reproduced from Baker, 2005). The name given to this attribute is DIMENSIONLIST.

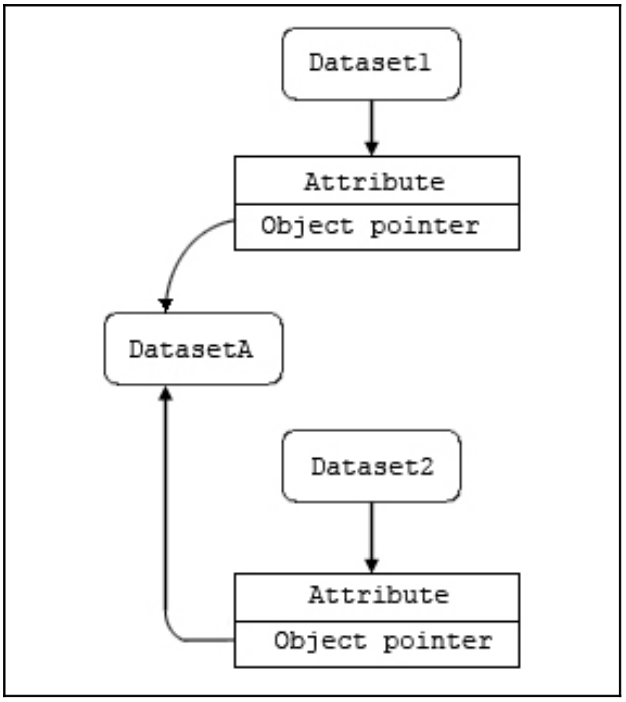


Figure 3.4: A shared HDF5 attribute and its associated dataset(s). DatasetA is a shared attribute of Dataset1. It is associated with Dataset1 by means of an object reference pointer attached as an attribute to Dataset1. Such an attribute can be shared among multiple datasets by means of additional object reference pointers attached to additional datasets (from Baker, 2005).

3.5 Numerical Types

The accepted binary representations (or "numerical" types) for the AVDC HDF4 and HDF5

[1] HDF5 does not have pre-defined SDS attributes.
[2] H4toh5.exe includes backward references to the original HDF4 file, but these are not incorporated in the AVDC H5 format.

type files are found in Table 3.4. These allowed types have been carefully chosen for compatibility and implementation in FORTRAN, C and HDF.

Table 3.4: Allowed binary representations ("numerical" types).

Numerical type	Comment
REAL	*HDF: 32 bit floating point numbers; FORTRAN: *4 real*
DOUBLE	*HDF 64-bit floating point numbers; FORTRAN: *8 real*
INTEGER	*HDF: 16-bit signed integers; FORTRAN: *2 integer*
LONG	*HDF: 32-bit signed integers; FORTRAN: *4 integer*
STRING	*character string*

4 Acronyms

AVDC	Aura Validation Data Center
ESA	European Space Agency
HDF	Hierarchical Data Format
NASA	National Aeronautical and Space Agency
NCSA	National Center for Supercomputing Applications
SDS	HDF4 Scientific Data Structure data model

5 Version History

2006/09/30 Addition of sub-sections 2.5-2.7 describing an attribute entry changes for VAR_AVG_TYPE, VIS_PLOT_TYPE, VIS_SCALE_TYPE, VIS_SCALE_MIN and VIS_SCALE_MAX.

6 References

F. Baker, HDF5 User's Guide, HDF5 Release 1.6.5, November 2005, UIUC / NCSA / HDF: http://hdf.ncsa.uiuc.edu/HDF5/doc/UG

B.R. Bojkov, De Mazière, M. and R. Koopman, Generic metadata guidelines on atmospheric and oceanographic datasets for the Envisat Calibration and Validation Project, Version 01R001, April 23, 2002.

NCSA, National Center for Supercomputing Applications, HDF Users Guide, Version 4.1r3,

May 1999: http://hdf.ncsa.uiuc.edu/UG41r3_htm