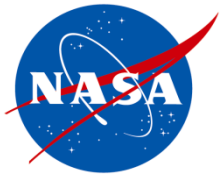

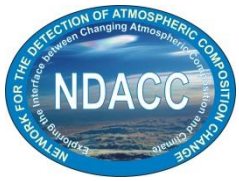

The Generic Earth Observation Metadata Standard (GEOMS) QA/QC Checks

Version 2.1
March 30, 2017

Author(s)
Ian Boyd, BC Scientific Consulting LLC

Co-author(s)
Ann Mari Fjaeraa, NILU
Martine De Mazière, BIRA-IASB
Sander Niemeijer, S[&]T
Jeannette Wild, NOAA
Bavo Langerock, BIRA-IASB
Roger Lin, NOAA
Ghassan Taha, NASA GSFC

Document Profile Information

Title:	The Generic Earth Observation Metadata Standard (GEOMS) QA/QC Checks
Issue	v2.1
Issue date:	March 30, 2017
Status:	Final
Audience:	Data providers
Reference(s):	The Generic Earth Observation Metadata Standard (GEOMS) [1]
Filename:	geoms_qa_checks_2.1.pdf
Distribution:	Public
Author(s):	Ian Boyd (BC Scientific Consulting LLC, USA, iboyd@astro.umass.edu)
Co-authors(s):	Ann Mari Fjaeraa (NILU, Norway, Ann.Mari.Fjaeraa@nilu.no) Martine De Mazière (BIRA-IASB, Belgium, Martine.DeMaziere@bira-iasb.oma.be) Sander Niemeijer (S[&]T, The Netherlands, niemeijer@stcorp.nl) Jeannette Wild (NOAA, USA, jeannette.wild@noaa.gov) Bavo Langerock (BIRA-IASB, Belgium, bavo.langerock@aeronomie.be) Roger Lin (NOAA, USA, roger.lin@noaa.gov) Ghassan Taha (NASA GSFC, USA, ghassan.taha-1@nasa.gov)

Contents

Document Profile Information 2
Introduction 4
1.0 Global Attribute Checks: 4
1.1 General 4
1.2 Originator Attributes 4
1.3 Dataset Attributes 4
1.4 File Attributes 5
2.0 Variable Attribute Checks: 6
2.1 General 6
2.2 VAR_NAME 6
2.3 VAR_SIZE 6
2.4 VAR_DEPEND 7
2.5 VAR_DATA_TYPE 7
2.6 VAR_UNITS 7
2.7 VAR_SI_CONVERSION 7
2.8 VAR_VALID_[MIN|MAX] 8
2.9 VAR_FILL_VALUE 8
2.10 HDF4 Pre-Defined Variable Attributes (units, _Fillvalue, valid_range) – Optional 8
3.0 Dataset Checks: 8
3.1 General 8
References 10

Introduction

This document outlines the QA checks performed on GEOMS compliant HDF and netCDF files when submitted to the ESA Validation Data Centre (EVDC) and Aura Validation Centre (AVDC) to ensure the files conform to the metadata standard. It can be considered as a supplement to the reference document The Generic Earth Observation Metadata Standard (GEOMS) [1].

This document is divided into 3 sections covering: Global Metadata Attribute checks, Variable Metadata Attribute checks, and checks on the datasets.

1.0 Global Attribute Checks:

1.1 General

1.1.1 All mandatory Global Attributes are present.
1.1.2 Redundant/historical Global Attributes are not present.
1.1.3 Each Global Attribute listed only once.
1.1.4 All entries use the ISO 646/US-ASCII character set.
1.1.5 Number of characters in the attribute entry are less than or equal to the maximum permitted.
1.1.6 No leading or trailing spaces in metadata values or sub-values except for free text attributes.
1.1.7 Attribute labels must be uppercase.
1.1.8 Non-standard global attributes are present (information only).
1.1.9 All sub-values are separated by semi-colons.

1.2 Originator Attributes

1.2.1 AFFILIATION value checked against the Table Attribute values file entries.
1.2.2 The COUNTRY sub-value in the ADDRESS entry checked against the Table Attribute values file entries.
1.2.3 The COUNTRY sub-value must be in upper-case or have initial capitalization.
1.2.4 The EMAIL address contains an '@' symbol within the entry i.e. it cannot be the first or last character.
1.2.5 The number of sub-values in the originator entry matches the expected number of sub-values.

1.3 Dataset Attributes

1.3.1 DATA_DISCIPLINE entry consists of 3 sub-values.
1.3.2 DATA_DISCIPLINE sub-values checked against the Table Attribute values file entries.
1.3.3 DATA_GROUP entry consists of 2 sub-values.
1.3.4 DATA_GROUP sub-values checked against the Table Attribute values file entries.
1.3.5 DATA_LOCATION entry consists of 1 value.

1.3.6 DATA_LOCATION checked against the Table Attribute values file entries.
1.3.7 DATA_SOURCE component of the DATA_SOURCE entry checked against the Table Attribute values file entries.
1.3.8 AFFILIATION component of the DATA_SOURCE entry checked against the Table Attribute values file entries.
1.3.9 3-character instrument identifier component of the DATA_SOURCE entry is a valid value.
1.3.10 DATA_SOURCE instrument identifier is '000' for sounding measurements using disposable instruments.
1.3.11 Construction of the DATA_SOURCE value is correct.
1.3.12 DATA_VARIABLES listing of variables match those present in the HDF file.
1.3.13 Each variable in DATA_VARIABLES is listed only once.
1.3.14 Number of DATA_VARIABLES matches the number of VAR_NAME entries.
1.3.15 The DATA_VARIABLES list has axis variables listed before their dependent datasets.
1.3.16 The DATA_START_DATE equals the lowest reported DATETIME value.
1.3.17 The DATA_STOP_DATE equals the highest reported DATETIME value.
1.3.18 The DATA_START_DATE and DATA_STOP_DATE values are written in the ISO 8601 format.
1.3.19 The DATA_START_DATE and DATA_STOP_DATE values are written in uppercase.
1.3.20 The DATA_START_DATE and DATA_STOP_DATE seconds value is between '00' and '59' (currently does not account for leap seconds).
1.3.21 DATA_FILE_VERSION is a 3-digit number greater than or equal to 001 and less than or equal to 999, in string form, with leading zeroes.
1.3.22 Check whether a DATA_TEMPLATE value is mandatory based on the DATA_SOURCE value.
1.3.23 If DATA_TEMPLATE value is present then entry checked against the Table Attribute values file entries.
1.3.24 DATA_QUALITY label is present if DATA_TEMPLATE label is present (otherwise it is optional).

1.4 File Attributes

1.4.1 Composition of the FILE_NAME value is correct based on the values that make up the file name.
1.4.2 The FILE_NAME extension is one of .hdf (HDF4), .h5 (HDF5), .nc (netCDF).
1.4.3 The actual file name corresponds to the FILE_NAME value.
1.4.4 The FILE_NAME value is written in lower case.
1.4.5 The FILE_GENERATION_DATE value is written in ISO 8601 format.
1.4.6 The FILE_GENERATION_DATE value is written in uppercase.
1.4.7 The FILE_GENERATION_DATE seconds value is between '00' and '59' (currently does not check for leap seconds).

1.4.8 FILE_ACCESS values or sub-values checked against the Table Attribute values file entries.
1.4.9 FILE_META_VERSION revision number sub-value must start with 04 or greater and be equal to or less than the current TAV file version.
1.4.10 FILE_META_VERSION tool name sub-value checked against the Table Attribute values file entries.
1.4.11 FILE_META_VERSION contains 2 sub-values only.

2.0 Variable Attribute Checks:

2.1 General

2.1.1 All mandatory reporting Variable Attributes are present for each dataset.
2.1.2 Check for redundant/obsolete/unknown variable attributes.
2.1.3 Each Variable Attribute listed only once for each dataset.
2.1.4 All entries use the ISO 646/US-ASCII character set.
2.1.5 Number of characters in the attribute entry are less than or equal to the maximum permitted.
2.1.6 One set of variable attributes for each dataset.
2.1.7 No leading or trailing spaces in metadata values or sub-values except for free text attributes.
2.1.8 GEOMS attribute labels must be uppercase.
2.1.9 All sub-values are separated by semi-colons.

2.2 VAR_NAME

2.2.1 HDF4 dataset name is not truncated (can happen with HDF4 library versions earlier than HDF4.2r2).
2.2.2 HDF or netCDF dataset name matches the VAR_NAME value.
2.2.3 VAR_NAME value contains no double underscores (used pre-GEOMS if variable mode not present).
2.2.4 VAR_NAME value matches corresponding DATA_VARIABLES value.
2.2.5 VAR_NAME value matches a combination of possible variable name, mode and descriptor values from the Table Attribute values file.

2.3 VAR_SIZE

2.3.1 The ordering of the VAR_SIZE values matches the array ordering of the dataset for multi-dimensional datasets.
2.3.2 The VAR_SIZE value(s) match the size of the respective dataset.
2.3.3 The VAR_SIZE value(s) are written to the file using STRING data type.
2.3.4 The number of dimensions for the dataset is less than or equal to 8.
2.3.5 The VAR_SIZE value(s) are written in the form of integers.

2.3.6 If the vertical dimension includes a '.BOUNDARIES' dataset (e.g. ALTITUDE.BOUNDARIES) then the fastest changing dimension must have a VAR_SIZE value of 2 and represent the lower and upper limits of each vertical level, except in the case where the ordering does not conform to the instrument template.

2.4 VAR_DEPEND

2.4.1 The ordering of the VAR_DEPEND values matches the array ordering of the dataset for multi-dimensional datasets.

2.4.2 The ordering of the VAR_DEPEND value(s) matches up with the VAR_SIZE value ordering.

2.4.3 The axis variable is self-referencing.

2.4.4 Non-axis variables are not self-referencing.

2.4.5 The axis variable is not dependent on another variable.

2.4.6 There is only one axis variable in the vertical dimension.

2.4.7 If the vertical dimension includes a '.BOUNDARIES' dataset (e.g. ALTITUDE.BOUNDARIES) then the fastest changing dimension must be labeled INDEPENDENT and represent the lower and upper limits of each vertical level, except in the case where the ordering does not conform to the instrument template.

2.4.8 VAR_SIZE value must be 1 if the VAR_DEPEND value is CONSTANT.

2.5 VAR_DATA_TYPE

2.5.1 Value must be one of BYTE (8-bit unsigned), SHORT (16-bit signed), INTEGER (32-bit signed), REAL (single precision), DOUBLE (double precision) or STRING.

2.5.2 Value must be DOUBLE if VAR_UNITS=MJD2K.

2.6 VAR_UNITS

2.6.1 VAR_UNITS value matches a combination of the permissible UNIT_PREFIX and VAR_UNITS values from the Table Attribute Values file list.

2.6.2 The value is written to the file using STRING data type.

2.6.3 If VAR_DATA_TYPE is STRING then the VAR_UNITS entry is [EMPTY].

2.6.4 VAR_UNITS must be MJD2K for any VAR_NAME that starts with DATETIME.

2.7 VAR_SI_CONVERSION

2.7.1 VAR_SI_CONVERSION values provide the correct formula used to convert the data specified in VAR_UNITS to the equivalent data in the corresponding base units.

2.7.2 If VAR_DATA_TYPE is REAL or DOUBLE then the corresponding VAR_SI_CONVERSION numeric values are in floating point notation, otherwise they are in integer notation.

2.7.3 All sub-values are written to the file using STRING data type.

2.7.4 If VAR_DATA_TYPE is STRING then the VAR_SI_CONVERSION entry is [EMPTY].

2.8 VAR_VALID_[MIN|MAX]

2.8.1 The data type of the VAR_VALID_[MIN|MAX] entry matches the data type of the respective dataset.

2.8.2 If VAR_DATA_TYPE is STRING then the VAR_VALID_[MIN|MAX] entry is [EMPTY].

2.8.3 The VAR_VALID_MIN value is less than or equal to the VAR_VALID_MAX value.

2.8.4 The VAR_VALID_MIN and VAR_VALID_MAX values for any latitude datasets are ≥ -90.0 and ≤ 90.0 respectively (GEOMS convention is + for northern hemisphere and - for southern hemisphere).

2.8.5 The VAR_VALID_MIN and VAR_VALID_MAX values for any longitude datasets are ≥ -180.0 and ≤ 180.0 respectively (GEOMS convention is + for east of Prime Meridian and - for west of Prime Meridian).

2.8.6 The VAR_VALID_MIN and VAR_VALID_MAX values for any zenith angle datasets are ≥ 0.0 and ≤ 180.0 respectively.

2.8.7 The VAR_VALID_MIN and VAR_VALID_MAX values for any azimuth angle datasets are ≥ 0.0 and ≤ 360.0 respectively (GEOMS convention is 0.0 for north, 90.0 for east and so on).

2.8.8 The VAR_VALID_MIN and VAR_VALID_MAX values for any wind direction datasets are 0.0 and 360.0 respectively (WMO convention is 0.0 for calm, 90.0 for east and 360.0 for north).

2.9 VAR_FILL_VALUE

2.9.1 The data type of the VAR_FILL_VALUE entry matches the data type of the respective dataset.

2.9.2 If VAR_DATA_TYPE is STRING then the VAR_FILL_VALUE entry is [EMPTY].

2.9.3 If the VAR_FILL_VALUE falls within the VAR_VALID_MIN and VAR_VALID_MAX values then it is defined as a default value (information only).

2.10 HDF4 Pre-Defined Variable Attributes (units, _Fillvalue, valid_range) – Optional

2.10.1 The data type of the pre-defined variable attribute entry matches the data type of the respective dataset.

2.10.2 Values match the corresponding GEOMS Variable Attribute values.

2.10.3 Any pre-defined attributes not listed above are not permitted.

2.10.4 Scaling or offset values are not permitted.

3.0 Dataset Checks:

3.1 General

3.1.1 All entries use the ISO 646/US-ASCII character set.

3.1.2 The array ordering of the dataset matches the given VAR_DEPEND and VAR_SIZE ordering.
3.1.3 The array size of the dataset matches that given in VAR_SIZE.
3.1.4 The array data type matches the given VAR_DATA_TYPE.
3.1.5 String datasets must be left justified.
3.1.6 All numeric data values fall within the VAR_VALID_MIN and VAR_VALID_MAX values.
3.1.7 All numeric data values fall within the range defined by the VAR_DATA_TYPE.
3.1.8 The DATETIME dataset does not contain fill values.
3.1.9 DATETIME values are in chronological order, and there are no duplicate values.
3.1.10 Datasets with VAR_UNITS=MJD2K are written to the file in MJD2K format (not ISO 8601).
3.1.11 Dataset does not contain NaN values.
3.1.12 That the reference vertical axis values are monotonically increasing or decreasing.
3.1.13 If the vertical dimension includes a '.BOUNDARIES' dataset then the order of values in the INDEPENDENT dimension are the same as the vertical axis corresponding to the .BOUNDARIES dataset e.g. from low to high if the values of the corresponding dataset are monotonically increasing.
3.1.14 If the vertical dimension includes a '.BOUNDARIES' dataset then the order of values in the vertical dimension are the same as the vertical axis corresponding to the .BOUNDARIES dataset.
3.1.15 If a LATITUDE dataset is provided then a LONGITUDE dataset must also be present, and vice versa. This also applies to related datasets e.g. LATITUDE.INSTRUMENT must be accompanied by LONGITUDE.INSTRUMENT.
3.1.16 If a WIND.DIRECTION (or related) dataset is provided then a WIND.SPEED (or related) dataset must also be present, and vice versa.
3.1.17 The VAR_SIZE values of any related datasets referred to in 3.1.15 and 3.1.16 must be the same.
3.1.18 If a WIND.SPEED value is 0.0 then the corresponding WIND.DIRECTION value must be 0.0 (WMO definition), and vice versa.
3.1.19 If a WIND.SPEED value is not 0.0 then the corresponding WIND.DIRECTION value cannot be 0.0 (WMO definition), and vice versa.
3.1.20 For any azimuth angle datasets the values must be ≥ 0.0 and < 360.0 degrees (GEOMS convention is 0.0 for north, 90.0 for east and so on).

References

[1] Retscher, C., De Mazière, M., Meijer, Y., Vik, A. F., Boyd, I. S., Niemeijer, S., Koopman, R. M., Bojkov, B., The Generic Earth Observation Metadata Standard (GEOMS), (2011): http://avdc.gsfc.nasa.gov/PDF/GEOMS/geoms-1.0.pdf