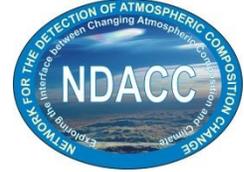




**evdc**  
esa eo validation data centre



**esa**



# The Generic Earth Observation Metadata Standard (GEOMS) Guidelines and Conventions

**Version 2.1**  
**April 17, 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) Guidelines and Conventions
Issue	v2.1
Issue date:	April 17, 2017
Status:	Final
Audience:	Data Providers, Data Archive Centers
Reference(s):	<a href="#">The Generic Earth Observation Metadata Standard (GEOMS)</a>
Filename:	geoms_guidelines_conventions_2.1.pdf
Distribution:	Public
Author(s):	Ian Boyd (BC Scientific Consulting LLC, USA, <a href="mailto:iboyd@astro.umass.edu">iboyd@astro.umass.edu</a> )
Co-authors(s):	Ann Mari Fjaeraa (NILU, Norway, <a href="mailto:Ann.Mari.Fjaeraa@nilu.no">Ann.Mari.Fjaeraa@nilu.no</a> ) Martine De Mazière (BIRA-IASB, Belgium, <a href="mailto:Martine.DeMaziere@bira-iasb.oma.be">Martine.DeMaziere@bira-iasb.oma.be</a> ) Sander Niemeijer (S[&]T, The Netherlands, <a href="mailto:niemeijer@stcorp.nl">niemeijer@stcorp.nl</a> ) Jeannette Wild (NOAA, USA, <a href="mailto:jeannette.wild@noaa.gov">jeannette.wild@noaa.gov</a> ) Bavo Langerock (BIRA-IASB, Belgium, <a href="mailto:bavo.langerock@aeronomie.be">bavo.langerock@aeronomie.be</a> ) Roger Lin (NOAA, USA, <a href="mailto:roger.lin@noaa.gov">roger.lin@noaa.gov</a> ) Ghassan Taha (NASA GSFC, USA, <a href="mailto:ghassan.taha-1@nasa.gov">ghassan.taha-1@nasa.gov</a> )

## Contents

Document Profile Information .....	2
Introduction.....	4
1.0 Incrementing the DATA_FILE_VERSION .....	4
2.0 Selecting the Correct File Version in case of Overlap .....	5
3.0 Differentiating between near-simultaneous Soundings or multiple Sondes on the same platform...	6
4.0 Reporting Conventions for Country Metadata Values and Latitude, Longitude, Azimuth, Wind Speed, Wind Direction and Transformation datasets .....	7
4.1 Country .....	7
4.2 Latitude and Longitude .....	7
4.3 Azimuth.....	7
4.4 Wind Speed and Direction .....	7
4.5 Transformation Dataset Ordering (e.g. Averaging Kernels) .....	8
5.0 Additional Template guidelines .....	8
References.....	9

# Introduction

This document describes guidelines and reporting conventions to supplement metadata and data requirements laid out in [The Generic Earth Observation Metadata Standard \(GEOMS\) \[1\]](#) document.

It is divided into 5 sections covering guidelines on [Incrementing the DATA\\_FILE\\_VERSION](#); [Selecting the correct file version in case of overlap](#); [Differentiating between near-simultaneous Soundings or multiple Sondes on the same platform](#), as well as [Reporting conventions for Country metadata values and Latitude, Longitude, Azimuth, Wind Speed, Wind Direction and Transformation datasets](#), and [Additional Template guidelines](#).

## 1.0 Incrementing the DATA\_FILE\_VERSION

The [GEOMS document \[1\]](#) describes the DATA\_FILE\_VERSION attribute as follows:

**The global attribute DATA\_FILE\_VERSION specifies the version of the data. It 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 shall be incremented by 1.**

**Type: STRING**

**Entry: Single field**

**Format: [ddd]**

**Example 1: DATA\_FILE\_VERSION = 003**

This section provides additional guidelines regarding usage of this variable.

1. A unique DATA\_FILE\_VERSION identifier should be able to describe a set of data that can be treated by the user of that data as a single continuous time-series. For example, if a PI is submitting preliminary data within a few days of the measurement (RD or rapid delivery) then that dataset might have a DATA\_FILE\_VERSION of 001. The final processed version of that same data would have a DATA\_FILE\_VERSION of 002.
2. The final consolidated version should be higher than the RD version.
3. If a PI submits the same measurements to another GEOMS affiliated database, then the current DATA\_FILE\_VERSION should be retained, even if it is not in sequential order on the second database – with the proviso that the DATA\_FILE\_VERSION value must be equal to or greater than that already submitted to either of the databases. This is to facilitate data exchange and categorizing of measurements between the affiliated data centers. It also means that the PI does not create additional versions of the same measurement with different DATA\_FILE\_VERSION values.
4. If a GEOMS template is modified and data has already been submitted to the data center, any measurements submitted with the new template should have an incremented DATA\_FILE\_VERSION. If submitting data for the first time then providers would still typically

start at 001. It is also recommended that previously submitted data be resubmitted using the new template.

5. Some data centers keep all submitted files on-line, while other data centers keep only the latest version on-line, with older versions retained off-line. The most recent version(s) of an on-line dataset can include one or more DATA\_FILE\_VERSION values. For example, if the PI is submitting both RD and final processed data, or if the PI has incremented the DATA\_FILE\_VERSION because of an instrument or algorithm change, and has not resubmitted older data.
6. Data centers will check for and reject files that have filenames matching those already on the database. In the case where there is any overlap in Data Start and Stop times refer to the section [Selecting the Correct File Version in case of Overlap](#).

## 2.0 Selecting the Correct File Version in case of Overlap

A conflict exists between two products if there is any overlap between the data start/stop date interval of the products, and the products contain data for the same instrument (i.e. DATA\_DISCIPLINE, DATA\_SOURCE, AFFILIATION and DATA\_LOCATION values are the same).

A Data User or Data Center never wants to have the same measurement twice. Even if data for just a single measurement sample is present in two products, an algorithm is required to determine which product has preference over the other (since there is no mechanism to discard only samples within a product).

Note that the case where the DATA\_STOP\_DATE of one product equals exactly the DATA\_START\_DATE of another product can sometimes be considered an overlap conflict and sometimes not. It is *not* considered an overlap if the DATA\_START\_DATE and DATA\_STOP\_DATE are based on the sample time *ranges* where DATA\_START\_DATE is taken from the first DATETIME.START value and DATA\_STOP\_DATE is taken from the last DATETIME.STOP value. However, if the DATA\_START\_DATE and DATA\_STOP\_DATE are based on sample time *points* (when e.g. DATA\_START\_DATE is taken from the first value of DATETIME and DATA\_STOP\_DATE is taken from the last value of DATETIME) then equality of DATA\_STOP\_DATE and DATA\_START\_DATE *is* considered an overlap (and a conflict).

For balloon soundings, it is possible for multiple instruments of the same type to be mounted on the same platform, or for soundings to have overlapping times when launched at the same location. To eliminate conflicts for these situations, a special rule applies for the use of DATA\_SOURCE. Please refer to the section [Differentiating between near-simultaneous Soundings or Multiple Sondes on the same Platform](#).

Once two products are considered to be in conflict the following algorithm, in the given order, can be applied to choose between the two:

1. If the filenames of the two products are exactly the same i.e. the same DATA\_DISCIPLINE, DATA\_SOURCE, AFFILIATION, 3-digit instrument identifier, DATA\_LOCATION, DATA\_START\_DATE, DATA\_STOP\_DATE, DATA\_FILE\_VERSION, then the Data Center will reject

the most recently submitted product, and notify the Data Originator (DO) or Principal Investigator (PI).

2. If it is determined that one of the products is Rapid Delivery (RD) and the other is not then the non-RD product has preference, even if the DATA\_FILE\_VERSION of the RD product is higher. To determine if a file contains RD measurements the Global Attribute DATA\_QUALITY must be checked for the term 'RD' at the beginning of the entry.
3. If one product has a higher DATA\_FILE\_VERSION, then that product has preference.
4. If one product has a FILE\_GENERATION\_DATE that is higher, then that product has preference (note that this can be overruled by a previous step in the algorithm).
5. If both also have the same FILE\_GENERATION\_DATE then the products should be exactly equal. If this is not the case then the Data Center or User must notify the PI or DO.

### 3.0 Differentiating between near-simultaneous Soundings or multiple Sondes on the same platform

During intensive intercomparison or validation balloon campaigns it is possible that multiple sondes of the same type are launched simultaneously or near-simultaneously from the same location. It is therefore important to be able to differentiate between the measurements in the event that the start and stop times of the files are the same or overlap, but otherwise all other attributes making up the filename are the same.

The [GEOMS document](#) [1] states the following with regard to 'consumables' and the DATA\_SOURCE 3-digit instrument identifier:

**A particular case exists for instruments that are used as "consumables", for example weather balloons that are often lost after the sounding flight. In such cases a unique identifier may be useless. The identifier 000 is therefore reserved for the non-unique case. A laboratory may re-use this particular identifier any number of times.**

For the above-defined scenario, the guidelines are expanded as follows:

1. To account for the case of overlapping or concurrent soundings using single-use or consumable instruments, the files for each sounding will be differentiated by incrementing the 3-digit instrument identifier in the DATA\_SOURCE, starting at '000'. All other attributes making up the filename must be the same, except possibly DATA\_START\_DATE and DATA\_STOP\_DATE.
2. If the DATA\_FILE\_VERSION is incremented this indicates a different version of the submitted data and the rules regarding [Selecting Version of GEOMS Product in case of overlap](#) must also be considered.

## 4.0 Reporting Conventions for Country Metadata Values and Latitude, Longitude, Azimuth, Wind Speed, Wind Direction and Transformation datasets

### 4.1 Country

- 4.1.1 The Country is reported as the third sub-value of the [PI|DO|DS]\_ADDRESS global attributes. It is an ASCII mapped version of the [ISO 3166-1 short name](#) and can be reported in full uppercase form or with initial capitalization e.g. 'Réunion' is reported as either 'Reunion' or 'REUNION' (without the acute accent above the 'é').
- 4.1.2 If using initial capitalization, words such as 'and', 'the', 'of' and 'd' remain lower case e.g. 'Cote d'Ivoire'.

### 4.2 Latitude and Longitude

- 4.2.1 Latitude is defined as having positive values for the northern hemisphere and negative values for the southern hemisphere, with a full range of -90.0 to +90.0 degrees.
- 4.2.2 Longitude is defined as having negative values west of Prime Meridian and positive values east of Prime Meridian, with a full range of -180.0 to +180.0 degrees.
- 4.2.3 If one of Latitude or Longitude is reported then it is mandatory to report the other.

### 4.3 Azimuth

- 4.3.1 Azimuth is defined using the 360 degrees of a compass, starting from north and increasing clockwise (0.0 for north; 90.0 for east and so on), with a full range of 0.0 to 360.0 degrees.
- 4.3.2 Values must be greater than or equal to 0.0 degrees and less than 360.0 degrees.

### 4.4 Wind Speed and Direction

- 4.4.1 Wind direction uses the WMO reporting convention. It is defined as the direction from which the wind is blowing, relative to true north. The direction is specified using the 360 degrees of a compass, starting from north and increasing clockwise. North is represented as 360.0 degrees as 0.0 degrees is used to represent calm. The full range is 0.0 to 360.0 degrees.
- 4.4.2 If wind speed is reported as 0.0 m/s, this indicates calm conditions, as defined by the sensitivity of the measuring device, and the wind direction must then be reported as 0.0 degrees, and vice versa.
- 4.4.3 If wind speed is not equal to 0.0 m/s then the direction cannot be 0.0 degrees and vice versa.

- 4.4.4 If one of wind speed or direction is reported then it is mandatory to report the other.

## 4.5 Transformation Dataset Ordering (e.g. Averaging Kernels)

A dataset describing a mathematical transformation has a source and target space dimension: e.g. a 2D averaging kernel matrix on an altitude grid acts on a concentration profile and returns a concentration profile and, from the GEOMS guidelines, VAR\_DEPEND is ALTITUDE;ALTITUDE. But some ambiguity remains to interpret which ALTITUDE dimension corresponds to the source or kernel space.

GEOMS puts the source dimension as the last dimension in VAR\_DEPEND. Hence, following the GEOMS guidelines, the source dimension (or the column index in standard matrix representation) is the fastest running dimension and, for an averaging kernel matrix, the source dimension corresponds to the 'kernel dimension'.

## 5.0 Additional Template guidelines

The [GEOMS document](#) [1] describes the use of templates for reporting as follows:

**Some GEOMS compliant measurements must be reported according to guidelines as set in GEOMS templates. In general, a GEOMS template states what metadata attributes and data variables must be reported. The general guidelines set above and throughout this document are valid for all templates and all rules must be followed. Templates cannot be more loose on general rules, but only more restrictive in the usage of metadata attributes and data variables. Templates may e.g. rule the usage of generally optional, but mandatory variable attributes if used in the GEOMS-TE context.**

In addition, templates describe the reporting of each data variable as being 'mandatory' i.e. it must be present in the GEOMS file; 'optional' i.e. it is up to the originator of the data whether or not to report the data variable and; 'conditional mandatory' i.e. the data variable must be reported if the described condition is met.

GEOMS also expands on the designation of 'conditional mandatory' to allow for the reporting of a data variable even if the condition is not met i.e. it also encompasses the definition for 'optional' reporting.

## 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>