

Appendix 6: Metadata

URL: <<http://badc.nerc.ac.uk/help/metadata/>>



Metadata

Contents

1. [What are metadata? Why are they essential?](#)
2. [Elements of metadata attached to specific types of data](#)
 - [2.1 Metadata for tables of numbers \(observations or model output\)](#)
 - [2.2 Metadata for images of the Earth surface](#)
 - [2.3 Metadata for software](#)
3. [Metadata standards and data formats](#)
 - [3.1 Metadata in NASA Ames format](#)
 - [3.2 Metadata in NetCDF and the CF standard](#)
4. [Complementary metadata](#)

Please address your comments on or suggested contributions to this page to the [BADC](#).

1. What are metadata? Why are they essential?

The term *metadata* encompasses all the information necessary to interpret, understand and use a given dataset. *Discovery metadata* more particularly apply to information (keywords) that can be used to identify and locate the data that meet the user's requirements (*via* a Web browser, a Web based catalogue, etc). *Detailed metadata* include the additional information necessary for a user to work with the data without reference back to the data provider (although one element of the detailed metadata may be the data provider's contact!).

Metadata pertaining to observational data, for example, include details about how (with which instrument or technique), when, where and with which accuracy (or error bars) the data have been collected, by whom (including affiliation and contact address or telephone number) and in the framework of which research project. In the case of processed data, the nature of the initial raw data and the derivation process must be stated. The nature and units of the recorded variables are of course essential, as well as the grid or the reference system. Metadata pertaining to model output should include the name of the model, the conditions of the calculation, the type of constraint applied, the length of the integration, the nature of the

Appendix 6: Metadata

URL: <<http://badc.nerc.ac.uk/help/metadata/>>

output, the geographical domain over which the output is defined (when applicable), etc. Specific conditions applying to the model or the experiment may be mentioned. Metadata also obviously include information on the format in which the data are stored, the order of the variables, etc, to allow potential users to read them. Metadata pertaining to software models include the key points of the theory on which the model is based, the techniques and computational language used, references, etc.

Metadata relative to a specific data set can be provided as a separate document or as a piece of the data set itself. For digital data sets, this means that the metadata can sit in separate files (for example text files) or be integrated into the data file(s), as a header or at specified locations in the file. Some data formats provide room and rules for metadata (see [Section 3](#)).

As far as possible, metadata of data held at the BADC follow the guidelines laid below. Data providers are encouraged to comply with the [BADC implementation of the Climate and Forecast \(CF\) Metadata Convention](#) (see also [Section 3](#)).

2. Elements of metadata attached to specific types of data

The following subsections list the minimal information that should ideally accompany certain types of data commonly archived for the use of atmospheric scientists.

2.1 Metadata for tables of numbers (observations or model output)

Metadata should include the following overall information. Some information in this list may be applicable in specific cases only.

- **Information about the experiment.**
 - Date when experiment or model simulation started.
 - Site or trajectory bounding box or domain limits.
 - Platform, instrumentation.
 - Model name.
- **Information about the experimenter(s).**
 - Names, affiliation, contact address including e-mail, telephone number.
 - Research programme name, research project code.
- **Information about the independent variables (usually spatio-temporal grid).**
 - Names, units, domain of definition of independent variables.
 - Interval values when appropriate.
- **Information about the data, including processing level.**
 - Version number.
 - Date of last revision.
 - Processing level (nature of raw data, derivation method).
 - Nature, name, units, scaling factors, accuracy of dependent variables.

Appendix 6: Metadata

URL: <<http://badc.nerc.ac.uk/help/metadata/>>

- **Information about data storage.**
 - Number of files of the entire dataset.
 - File sizes.
 - File number of current file.
- **Information about data format.**
 - Archive structure.
 - File structure.
 - Number of lines in file header if any.
 - Record structure.
- **Additional information.**
 - May include particular conditions of experiment or model run, model boundary conditions, article reference, source of further information, or other comments.

2.2 Metadata for images of the Earth surface

Elements of metadata of maps and (photographic, satellite,...) images of the Earth surface should include the following.

- **Information about the picture.**
 - Date when picture was taken.
 - Date of last revision, if any.
 - Geographical resolution and coverage.
 - Orientation. Platform, technique used, wavelength channel.
 - Picture resolution (real size corresponding to pixel).
- **Information about the experimenter(s).**
 - Names, affiliation, contact address including e-mail, telephone number.
 - Research programme name, research project code.
- **Information about picture storage.**
 - Number of files of the entire images set.
 - File sizes.
 - File number and name of image file.
- **Additional information.**
 - Photographic treatment.
 - Experiment associated with current image.
 - Any relevant information regarding the conditions when the picture was taken (e.g. meteorological conditions).
 - Any relevant information on the way the map was produced or the image derived.

2.3 Metadata for software

Metadata pertaining to a model should include the following.

- **Information on the model**
 - Brief description of model general aim.

Appendix 6: Metadata

URL: <<http://badc.nerc.ac.uk/help/metadata/>>

Model structure.

Physical processes involved, including equation set.

Parameterisations.

Algorithmic implementation techniques used.

Spatio-temporal coverage when applying.

Boundary conditions, including reference(s).

Initial conditions, including reference(s).

Program language.

Input nature and format.

Output nature and format.

Summary of model validation, or appropriate reference(s).

Summary of results from former studies conducted with the model, or appropriate reference(s).

- **Information on the author(s)**

Names, affiliation, contact address including e-mail, telephone number.

Research programme name, research project code.

- **Information on how to run the model**

Platforms, operation language, script.

Input files.

N.B. Metadata relative to software are commonly included as comments, either in the top section of the source file, or at various places of the code.

3. Metadata standards and data formats

Since the evaluation of information relevance may widely vary with individuals, some metadata standards have been and are still currently being developed with the aim of uniformising metadata presentation. The other advantage of metadata standards is that they ensure the transmission of the information contained in the metadata (and hence the ability to use the data), in some predefined generic way, to remote and future users, provided that the latter will know the adopted conventions. Which in turn requires the existence, maintenance and transmission of manuals describing the set of conventions relevant to a particular metadata standard some kind of *metametadata*.

Since a crucial section of the metadata pertains to the data format, different metadata standards have been developed in conjunction with the various data formats. (To know about the formats supported by the BADC, please refer to the BADC [Formats Welcome Page](#)). Existing data format standards, and metadata standards alike, are based both on the specific needs of confined scientific communities and on habits already in use within these communities. All of them regularly undergo updates and are susceptible of further evolution. In geosciences and among disciplines where 2-dimensional Earth surface reference systems play an important role (like archæology), the most popular data formats seem to belong to the [GIS](#) family (*Geographic Information Systems*). In the atmospheric research community, however, the third spatial (vertical) dimension obviously plays a crucial role, and so does

Appendix 6: Metadata

URL: <<http://badc.nerc.ac.uk/help/metadata/>>

time. Sections [3.1](#) and [3.2](#) below respectively give a brief outline of two formats widely used in the atmospheric sphere, namely the *NASA Ames Format for Data Exchange*, applying to data coded in ASCII, and *NetCDF* (network Common Data Form), applying to data coded in binary language and hence better adapted to voluminous data sets such as 3- or 4-dimensional fields, satellite data, etc. Both data formats include some metadata rules.

Standard rules can be mandatory, conditional or optional. They apply to three aspects of the metadata:

- **Content.** Which elements of information must/should/may be recorded.

Section 2 above is an attempt to answer this issue.

- **Vocabulary.** Standard terms that must/should/may be used to describe the elements of information (i.e. allowed units, allowed variable names, etc.).

A standard thesaurus is currently developed at the BADC to address this issue, based on the [Climate and Forecast \(CF\) Metadata Convention](#), the [Directory Interchange Format \(DIF\)](#) of the [Global Change Master Directory \(GCMD\)](#) and the specific needs of the BADC and of atmospheric scientists.

Meanwhile, data providers are strongly encouraged to comply with the [BADC implementation of the CF convention](#) (see [Section 3.2](#) below), even when using a data format different than NetCDF (such as NASA Ames).

- **Layout.** Order and syntax of the recorded elements (i.e. the format).

This issue is addressed by each specific data format. It is essential to allow the retrieval of the information by a piece of software.

3.1 Metadata in NASA Ames format

The [NASA Ames Format for Data Exchange](#) has been developed by S. Gaines and S. Hippskind at the NASA Ames Laboratory, for the benefit of instrument scientists operating atmospheric probe apparatus onboard balloons and aircrafts, and its straightforwardness and portability serve this purpose perfectly. It is in principle able to deal with 3- and 4-dimensional data sets, although the data layout within a file, which shows its original aim (i.e. the storage of time series), does not optimise the representation of fields on a 3-D or 4-D gridded domain. NASA Ames formatted data are coded in ASCII, which presents the noticeable advantage of being directly readable by (English speaking) humans, but the drawback of producing cumbersome files, which again is not optimal for 3-D or 4-D variables. Each NASA Ames file is divided into a header and a body, the latter containing the data, the former the metadata. The required metadata include both discovery and detailed metadata.

NASA Ames rules include some statements about the metadata content. Any additional

Appendix 6: Metadata

URL: <<http://badc.nerc.ac.uk/help/metadata/>>

information (for example, elements listed in [Section 2.1](#) that would not fit into the provided rules) can still be inserted in dedicated comment lines at the end of the header. The metadata layout is strictly defined in the NASA Ames format, but for the comment lines, which are loosely constrained. A complete description of the NASA Ames data and metadata format (including content and layout rules) is available from the BADC [NASA Ames Format Page](#).

The NASA Ames format makes no statement on any mandatory or suggested vocabulary. As mentioned earlier, data providers using NASA Ames are strongly encouraged to follow the [BADC guidelines on CF conventions](#) (see also [Section 3.2](#) below).

3.2 Metadata in NetCDF and the CF standard

[NetCDF](#) is the binary data format underlying the [Network Common Data Form](#) supported by [Unidata](#). It allows the user to insert metadata in the data files.

The [NetCDF Climate and Forecast \(CF\) Metadata Convention](#) has developed a standard dealing mainly with vocabulary rules. Although this standard was developed with the NetCDF format in mind, it can be applied to any set of geophysical data, and probably extended to cover a much broader range of disciplines as well.

With the aim of providing a consistent way of describing atmospheric data sets, the BADC has developed its own implementation of CF metadata rules. If you are about to submit metadata to the BADC, whether you use NetCDF or not, please refer to the [BADC implementation of the CF Convention](#).

4. Complementary metadata

Any additional documentation on recorded data or images, whether pertaining to a single data file or a whole dataset, that would not find its place into the structures described above (because it does not fall into any described category or because it is too voluminous) may be attached to the data, for example in the form of text files. These documents may for example include technique description, possible use of the data, study conclusions, etc.

Please address your comments on or suggested contributions to this page to the [BADC](#).