MOPITT Data File Format Description

The S2K MOPITT data files are in the ICARTT data file format. This is a recent modification of the NASA/Ames format

This format description and examples have been excerpted from "ICARTT Data Format".

Eric Williams, AL/NOAA; Jim Crawford and Ali Aknan, LaRC/NASA; Hans Schlager, DLR; (January 2006)

[http://www-air.larc.nasa.gov/missions/etc/IcarttDataFormat.htm]

3. Recommended File Format Specification for ICARTT Time-series Data Files

A. Structure

We recommend that, whenever possible, ICARTT time series data files conform to the following Ames file format:

FFI = 1001; one real, unbounded independent variable; primary variables are real; no auxiliary variables; independent and primary variables are recorded in the same record.

What this means in English is that there is one time (independent) variable and that all other data depend on that variable. Any number of other variables can be defined, but they all depend on the one. In the typical case the fundamental variable is the start time of the measurement and others can be defined as in the following example, where the variable names refer to columns in the data file:

    start time
    stop time
    mid-point time
    latitude
    longitude
    altitude/elevation
    data variable1
    variable1 uncertainty
    data variable2
    variable2 uncertainty
    <etc.>

This format accounts for most time series data measured anytime, over any arbitrary integration period, and at any place on or above the planet (within reason for air quality data). Obviously, the format can be condensed. For example, if measurements are reported as 1 second intervals, then stop time and mid-point time need not be included as data columns provided all time intervals in the measurement period are accounted for by inclusion of the missing data flag(s). Similarly, if the measurements are made at a fixed location then latitude, longitude, and elevation are fixed and

these data would be included in the header information (see below). As pointed out above, if the location data (latitude, etc.) are included in a separate file, then these columns can be excluded provided the location data file name is included in the header information for the data file. Similarly, if uncertainty is defined as some function that is the same for all data points then that function can be included in the header information and the user can then calculate uncertainties. Variations in the way the format is used, based on the needs of the data provider, are accounted for in the file header information. As an example, some PIs may wish to report the END time of the measurement period as the independent variable. The ICARTT format allows this provided that the time variable is clearly labeled as such (e.g., End_UTC) and that additional information describing this (non-standard) situation be provided in the Normal Comments section of the file header. If the data periods are not of a constant duration, then the start time and mid-point time of each period must be included as an additional column and the Data Interval value set to 0 (see below). The header specifications are described below.

B. File header information

The basic structure of the ICARTT file header is similar to the Ames exchange format. For the ICARTT study we recommend some additional information that will be included in the comments sections. The most general header is shown below as an example; more specialized headers will be described as modifications to the general form. Different items of information in the same record (same line) are shown below as separated by a semicolon – in the actual file they are separated by a single space.

*
Number of lines in header; file format index: most files use 1001
*
PI name: last name, first name/initial
*
Organization/affiliation of PI
*
Data source description: e.g., instrument name; platform name; model name, etc.
*
Mission name: (for the ICARTT study, this will be ICARTT_ followed by your project; e.g., NEAQS, INTEX, etc.
*
File volume number; number of file volumes: These integer values are used when the data require more than one file per day. For data that require only one file these values are 1 1.
*
UTC date when data begin; UTC date of data reduction or revision.
*
Data Interval: This value describes the time spacing (in seconds) between consecutive data records. It is the (constant) interval between values of the independent variable. For 1 Hz data the Data Interval value is 1; for 1 minute data the value is 60; for 2 Hz the value is 0.5. All intervals less than 1 Hz (exception for 1 minute data) must be reported as Start_UTC, Stop_UTC, and Midpoint_UTC; and the Data Interval value in the header is set to 0.
*

Description or name of independent variable: This will be the name chosen for the start time or in some cases the mid-point time or end time of the data stream. It always refers to the number of seconds from the UTC start of the day.
    *
    Number of variables: Integer value showing the number of dependent variables (the total number of columns of data will be this value plus one).
    *
    Scale factors: This will be 1 for all variables, except where grossly inconvenient.
    *
    Missing data indicator: This will be –9999 (or -99999, etc.) for any missing data condition, except for the main time variable which is never missing.
    *
    Variable names and units: Name or description, including units, of data in that column. If the variable is unitless, enter "none" for its units. Each variable name and units are entered on one line, and separated by a comma or semicolon.
    *
    Number of SPECIAL comment lines: Integer value indicating the number of lines of special comments, NOT including this line.
    *
    Special comments: Notes of problems or special circumstances unique to this file. An example would be comments/problems associated with a particular flight.
    *
    Number of Normal comments (i.e., number of additional lines of SUPPORTING information): Integer value indicating the number of lines of additional information, NOT including this line.
    *
    Normal comments (SUPPORTING information): This is the place for investigators to more completely describe the data and measurement parameters. The supporting information structure is described below as a list of key word: value pairs. Specifically include here information on the platform used, the geo-location of data, measurement technique, and data revision comments. Note the non-optional information regarding uncertainty, the upper limit of detection (ULOD) and the lower limit of detection (LLOD) for each measured variable. The ULOD and LLOD are the values, in the same units as the measurements that correspond to the flags –7777 and –8888 within the data, respectively. The last line of this section should contain all the variable names on one line. The key words in this section are written in BOLD for clarity below. The actual file will not have special formatting codes. The key word must be typed followed by a colon then followed by your text (information). When more than one value (or information) is to be written on the same line, separate the values using a semicolon. For lines where information is not needed or applicable, simply enter N/A. The scanning program will look for these key words (case insensitive) when the file is submitted.

    PI_CONTACT_INFO: Phone number, mailing address, email address and/or fax number.
    PLATFORM: Platform or site information.
    LOCATION: including lat/lon/elev if applicable.
    ASSOCIATED_DATA: File names with associated data: location data, aircraft parameters, ship data, etc.
    INSTRUMENT_INFO: Instrument description, sampling technique and peculiarities, literature references, etc.
    DATA_INFO: Units and other information regarding data manipulation.

UNCERTAINTY: Uncertainty information, whether a constant value or function, if the uncertainty is not given as separate variables.

ULOD_FLAG: -7777 (Upper LOD flag, always -7's).

ULOD_VALUE: Upper LOD value (or function) corresponding to the -7777's flag in the data records.

LLOD_FLAG: -8888 (Lower LOD flag, always -8's).

LLOD_VALUE: Lower LOD value (or function) corresponding to the -8888's flag in the data records.

DM_CONTACT_INFO: Name, affiliation, phone number, mailing address, email address and/or fax number.

PROJECT_INFO: Study start & stop dates, web links, etc.

STIPULATIONS_ON_USE: (self explanatory)

OTHER_COMMENTS: Any other relevant information.

REVISION: R# (see filenames discussion above AND the 3 examples below);

R#: comments specific to this data revision. The revision numbers and the associated comments are cumulative in the data file. This is required in order to track the changes that have occurred to the data over time. Prepend the information to this section so that the latest revision number and comments always start this part of the header information. The latest revision data should correspond to the revision date on Line 7 of the main file header. Note that FIELD data files have revision LETTERS, not numbers. The file checking software will be looking for these "R#" to ensure compliance.

Indep_Var VarName_1 VarName_2 VarName_3 … VarName_n

The formula for the total number of lines in the header for FFI=1001 files:
14 + (# dependent variables, given in line 10) + (# special comments) + (# normal comments)

C. Example

Below is an example of time series data. Be aware that the automatic word-wrap feature in word processing programs gives the appearance that there are more lines of text than are really there. In these examples any continuation of lines from directly above has been indented for clarity.

EXAMPLE 1. All required data columns are shown explicitly.
File name: NOX_RHBrown_20040830_R0.ict

41 1001
Williams, Eric
Aeronomy Laboratory/NOAA
Nitric oxide and nitrogen dioxide mixing ratios from R/V Ronald H. Brown
ICARTT_NEAQS
1 1
2004 08 30 2004 12 25
0
Start_UTC, (number of seconds from 0000 UTC)
9
1 1 1 1 1 1 1 1 1
-9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999
Stop_UTC, seconds

Mid_UTC, seconds
DLat, degrees
DLon, degrees
Elev, m
NO, ppbv
NO, 1sig
NO2, ppbv
NO2,1sig
0
18
PI_CONTACT_INFO: Address: 325 Broadway, Boulder, CO 80305; email: eric@al.noaa.gov;
303-497-3226
PLATFORM: NOAA research vessel Ronald H. Brown
LOCATION: Latitude, longitude and elevation data is included in the data records
ASSOCIATED_DATA: N/A
INSTRUMENT_INFO: NO: chemiluminescence; NO2: narrow-band
photolysis/chemiluminescence
DATA_INFO: All data with the exception of the location data is in ppbv. All one-minute averages
contain at least 35 seconds of data, otherwise missing.
UNCERTAINTY: included in the data records as variables with a _1sig suffix
ULOD_FLAG: -7777
ULOD_VALUE: N/A
LLOD_FLAG: -8888
LLOD_VALUE: N/A; N/A; N/A; N/A; N/A; 0.005; N/A; 0.025; N/A
DM_CONTACT_INFO: N/A
PROJECT_INFO: ICARTT study; 1 July-15 August 2004; Gulf of Maine and North Atlantic
Ocean
STIPULATIONS_ON_USE: Use of these data requires PRIOR OK from the PI
OTHER_COMMENTS: N/A
REVISION: R0
R0: No comments for this revision.

Start_UTC	Stop_UTC	Mid_UTC	DLat	DLon	Elev	NO_ppbv	NO_1sig	NO2_ppbv	NO2_1sig
43200	43259	43229	41.00000	-71.00000	15	0.555	0.033	2.220	0.291
43260	43319	43289	41.01234	-71.01234	15	10.333	0.522	31.000	0.375