

ICARTT File Format Standards V1.1

Status of this Memo

This RFC document describes the ICARTT file format standards, implementations and resources available.

Distribution of this memo is unlimited.

Change Explanation

Updates from V1 to V1.1:

Section 2.1.B. Location Information – Added clarification about precision of measurements. “It is recommended that the decimal latitude and longitude be reported at maximum instrument precision. For typical aviation GPS instruments, the latitude and longitude should be reported to at least five decimal places. Altitude is recommended to be reported in meters.”

Section 2.6. File Scanning Software – Corrected grammatical error in first sentence.

Copyright Notice

Copyright © United States Government as represented by the Administrator of the National Aeronautics and Space Administration. (2013). All Rights Reserved.

Abstract

The ICARTT file format standards were developed to fulfill the data management needs for the **International Consortium for Atmospheric Research on Transport and Transformation (ICARTT)** campaign in 2004. The ICARTT study consisted of eleven highly coordinated individual field experiments with over 300 government-agency and university participants from five countries, i.e., US, Canada, UK, Germany, and France. A common and simple-to-use data file format, ICARTT file format was established for this study to primarily facilitate data exchange and to promote collaborations among the science teams for achieving the ICARTT science objectives. The ICARTT file format is text-based and composed of a header section (metadata) with critical data description information (e.g., data source, uncertainties, contact information, and brief overview of measurement technique), and a data section. Although it was primarily designed for airborne data, the ICARTT format proved to be practical for other mobile and ground-based studies and various data types. Upon the success of the ICARTT study, the ICARTT file format has since been widely accepted in the atmospheric composition field study community and used in recent major airborne studies sponsored by NASA, NSF, NOAA and international partners.

Table of Contents

STATUS OF THIS MEMO	1
COPYRIGHT NOTICE	1
ABSTRACT	1
TABLE OF CONTENTS	2
1. INTRODUCTION -- ORIGIN OF THE ICARTT FILE FORMAT STANDARD	3
2. FILE FORMAT SPECIFICATIONS	4
2.1.A. TIME INFORMATION	5
2.1.B. LOCATION INFORMATION	5
2.1.C. MEASUREMENTS	5
2.2. FILE NAMES	6
2.3. FILE FORMAT SPECIFICATION FOR ICARTT TIME-SERIES DATA FILES	8
2.3.A. STRUCTURE	8
2.3.B. FILE HEADER INFORMATION	10
2.3.C. EXAMPLES	12
2.4. FILE FORMAT SPECIFICATION FOR ICARTT MULTI-DIMENSIONAL DATA FILES.....	15
2.4.A. STRUCTURE	15
2.4.B. EXAMPLES	15
2.5. FILE FORMATS FOR NON-STANDARD AIRBORNE DATA	18
2.6. FILE SCANNING SOFTWARE.....	19
3. REFERENCES	19
4. AUTHORS' ADDRESS.....	20
5. APPENDIX A - GLOSSARY	20

1 Introduction – Origin of the ICARTT file format standard

Since the early 1980s NASA and partner agencies have conducted over 30 major tropospheric airborne field campaigns to investigate atmospheric composition over a wide range of geographical regions. Compared to satellite data, airborne data provides a longer historical perspective, a more extensive suite of observed species/parameters, and higher spatial resolution both horizontally and vertically. Consequently, airborne observations are of unique value for the modeling community to assess its ability to predict future atmospheric composition and its impact on climate change and air quality issues. Furthermore, airborne observation can also be used effectively to develop and/or improve the a priori data used in satellite retrieval algorithms. Nevertheless, there are significant challenges in using airborne data for model assessment and validation. Among these is lack of uniform data format. The existing airborne measurement data are archived in various formats, which make the use and exchange of data difficult. Establishing standard data file protocols is one important step towards facilitating the data exchange between the scientific communities.

The **International Consortium for Atmospheric Research on Transport and Transformation (ICARTT)** field study was conducted in summer 2004, which consisted of eleven independent but highly coordinated field experiments, e.g., NASA INTEX-NA, NOAA NEAQS-ITCT, and EU ITOP. While each of these field studies had regionally focused science objectives (being sponsored by government agencies in five countries on both sides of the North Atlantic, i.e., US, Canada, UK, Germany, and France), collectively they all shared a common overarching scientific objective to examine the key processes related to the emissions of aerosol and ozone precursors and their chemical transformations and removal during transport to and over the North Atlantic. The ICARTT campaign involved several hundred participants and multiple airborne and shipboard platforms as well as ground-based components. A large volume of data was being generated by individual investigation groups within the corresponding field experiment. A common data file format, the ICARTT format, was created to accommodate data sharing among the science teams and to fulfill the data management needs for all phases of field study, i.e., field deployment, post deployment data processing and analysis and publications.

The ICARTT file format is a text-based, self-describing, and relatively simple-to-use file structure. The file format was built on two well-established airborne data formats: NASA Ames and GTE. Like its predecessors, the ICARTT file format is designed for handling airborne in-situ measurement data but having limited capability to accommodate data from airborne or ground-based remote sensing (e.g. lidar), ground-based measurements, and aspects of satellite data. The ICARTT format is composed of two sections: a header section (metadata) and a data section. The header section has the instructions for extracting data from the file and the critical information describing the data (e.g., data source, contact information, brief description of measurement technique, measurement uncertainties, and data revision comments) so that a user would have sufficient information to either make direct use of the data or contact the measurement PI to get further clarification on certain issues.

Because of the ICARTT field campaign, the ICARTT data file format was exposed to a broad range of airborne researchers. The success of the ICARTT campaign naturally led to even wider acceptance of the ICARTT file format in later airborne studies. For example, the ICARTT file format was adopted in NASA INTEX-B and NSF MILAGRO field campaigns in 2006. Even more recently, the international polar year POLARCAT field study used the ICARTT file format as the standard for the participating programs sponsored by NASA, NOAA, and international partners in France and Germany. The growing acceptance and wide use, especially in the airborne in-situ measurement community, has propelled the ICARTT data file format to be recognized as one of the standards for the airborne study community.

This document defines the ICARTT file format standards in section 2, which includes format specification, naming convention, header section specification, and applications to various data types. Several examples are provided here for further clarification. Also given here is a brief description of file scanning software that can be used to test data files for compliance with the ICARTT file format standards.

2 File Format Specifications

In large airborne field studies such as ICARTT 2004, there are many different types of data collected. Many data sets are simply straight time series with one or a number of parameters being measured sequentially (and simultaneously) in time. However, there are some data sets that are multi-dimensional in that observations taken by an instrument at a single point in time are spatially distributed.. An example is wind profiler data in which 30-minute averaged samples taken at some time period are binned into height information while at each height wind speed, wind direction, and temperature are supplied. Another more extreme example is the output from 3-dimensional models. Data such as these clearly cannot be represented as a single time series. Sections 2.1 - 2.2 below outline the ICARTT format for different types of data, with an emphasis on standard time-series types of data, which is typical from in-situ chemical measurements. Sections 2.3 - 2.4 are specific for standard time-series types of data. Section 2.5 offers guidance for non-standard time-series data.

The ICARTT file format has no restriction on the number of characters per line or on the number of characters per record. The file name is limited to 127 characters in length. Like its predecessors, the ICARTT format will also use the ASCII character set for file construction. The data section of the file will be comprised only of ASCII numeric characters including scientific notations, commas as delimiters, and spaces for the purpose of visual clarity (alignment) of data.

The end-of-line (EOL) character for text files differs on different operating systems. Many modern text utilities do handle and convert the EOL character automatically, and this problem to the vast majority of users is transparent and a non-issue. It is nonetheless a problem that some users will encounter, and data managers should be aware of it. There are many resources on the web (e.g. Wikipedia) discussing this issue in detail with remedies to overcome it. A quick fix can

be as simple as using the "ASCII mode" for ftp file transfer, or using a different (newer) text editor (e.g., WordPad), etc. The file scanning software mentioned (below) will automatically handle the EOL character when necessary.

2.1.A. Time information

Data files are required to report the start and stop time for each measurement. This is the only unambiguous way to represent measurement integration time. One exception is for data collected continuously at 1 Hz or less which may be represented by a single timestamp. Time is to be reported as seconds UTC from the start of the date on which measurements began. This date appears in both the file header and filename. The reported time should be monotonically increasing even when crossing over to a second day. For continuous measurements, the reported timeline must be unbroken between the first and last reported measurements. This is to be accomplished using missing data identifiers to account for data gaps due to calibration or other periods of instrument down time. Two important exceptions for timeline continuity are for data with irregular spacing and/or integration times and for satellite data which can experience significant gaps due to cloud interference, mode changes, etc. The satellite exception is in acknowledgement of the extremely large volumes of data acquired from these platforms (see Section 2.5 below).

2.1.B. Location information

All data points need to have an associated latitude, longitude, and altitude. Latitude and longitude should be reported in decimal degrees with south latitudes and west longitudes represented as negative numbers (i.e., no N, E, W, S identifiers). It is recommended that the decimal latitude and longitude be reported at maximum instrument precision. For typical aviation GPS instruments, the latitude and longitude should be reported to at least five decimal places. Altitude is recommended to be reported in meters. Altitudes must be explicitly defined since many types of altitude measurements are in use (pressure alt; GPS alt; geopotential alt; radar alt; etc.). Oftentimes, it is advantageous to report location (and other) information in an independent file (e.g., an aircraft parameter file). In that case, it is not necessary that this information be reported redundantly in the data files for each instrument on the platform. Instead, they may simply refer to the parameter file in the data file header. This option is specified below in section 2.3.B.

2.1.C. Measurements

In general, each file contains data of one parameter or species. Multiple variables per file are allowed only if all were measured on exactly the same time base, as, for example, by the same instrument. The numeric representation of a variable is defined by the units in which it was measured. The ICARTT format contains the provision for a data scaling factor. However, it is recommended that all scale factors be 1 unless it is grossly inconvenient to do so. If very large or very small numbers are required, then they can be represented with exponential notation, as in 1.01E9 or 5.23E-6.

i. Uncertainties

Measurement uncertainty is inherently associated with each measurement. The ICARTT data format requires reporting the TOTAL uncertainty to include all systematic and random effects. If the uncertainty estimates are available for each measurement period, the uncertainties can be tabulated as the next (and separate) column after the data column in the file. However, this requirement can be relaxed if the uncertainty data can be reproduced by information in the header of the file. For example, if all uncertainties can be calculated by a function that has any given data point as input, then the formula can be included as header information. It is imperative that the sigma confidence interval (e.g., 1 sigma or 2 sigma) should be reported with the uncertainty. Equally important, the units for the uncertainty must be explicitly reported in the file header. When absolute uncertainty is reported, the same unit should be used for the uncertainty as the associated measurement. For relative uncertainty, the value should be reported in percentage (e.g., 30% or 10%).

ii. Missing data

Missing data are just that – missing, i.e., instrument was not taking data due to calibration or instrument problem. Missing data are represented by negative numbers large enough to never be construed as actual data. For the ICARTT file format the value is -9999 (or -99999, etc.). On the other hand, data below (or above) the limit of detection (LOD) are not actually “missing” but do convey some information. While some investigators choose to tabulate all of their quantifiable data, including negative values for concentrations, others choose to show these data points as the values less than some quantifiable measurement limit. Similar treatment is also done for data with values greater than the upper LOD. These conditions are indicated by two additional missing data flags that are substituted for the missing data values. The flag for data values GREATER THAN some UPPER LOD (ULOD) is -7777 (or -77777, etc.), and the flag for data values LESS THAN some LOWER LOD (LLOD) is -8888 (or -88888, etc.). These flags (if used) and the values of the upper and lower LOD are documented at specific locations in the header file (see below). If LLOD or ULOD values vary from point to point, they should be given in a separate column of data.

iii. Data delimiter characters

Commas are used to delimit data fields within records (lines) of data in a file.

2.2. File names

Features of different file naming conventions have been adapted here. File names for the ICARTT data format, limited to 127 characters or less, are defined as follows:

dataID_locationID_YYYYMMDD[hh[mm[ss]]]_R#[_L#][_V#][_comments].ict,

Where the only allowed characters are: a-zA-Z0-9_.- (that is, upper case and lower case alphanumeric, underscore, period, and hyphen). All fields not in square brackets are required. Fields are described as follows:

dataID: short description of measured parameter/species, instrument, or model
(e.g., O3; RH; VOC; PTRMS; MM5)

locationID: short description of site, station, platform, laboratory or institute

YYYY: four-digit year

MM: two-digit month

DD: two-digit day

hh: optional two-digit hour

mm: optional two-digit minute

ss: optional two-digit second

R: revision number of data

L: optional launch number

V: optional volume number

comments: optional additional information

extension: ict file extension, always “ict”

The underscore is used ONLY to separate the different fields of the file name; it has special significance for file-checking software (see section 2.6). To separate characters within a field for readability, use lower and upper case letters. The use of the hyphen, though allowed, is discouraged since this character in file names may cause problems with some older operating systems and network software. The square brackets “[]” enclose optional parameters but are not shown in the file name. Dates and times in file names are always UTC. The date and time in the file name give the date/time at which the data within the file begin (data files), or date/time at which the image applies (image files). For aircraft and sonde data files, the date always refers to the UT date of launch.

The **dataID** is a short string of characters used to identify the parameters in the file. For files that contain one or two variables those variable names can be used in the file name. For files in which many variables are represented, it may be best to indicate in the file name a class of compounds (e.g., VOC; Photolysis Rates) or an abbreviation of the instrument used to make the measurements (e.g., PTRMS).

The **locationID** is used to identify the measurement platform, site, station, or source (laboratory or institute) of the information within a data file. Some examples could be: DC8, BAE146, RHBrown, GOME (satellite), IoS (Appledore Island site), ChebPt (Chebogue Point site), and others. It may be useful to have a standardized set of abbreviations used for a given field mission. These should be decided upon by the mission Science Team.

The **R** parameter is not optional in the ICARTT data format. One must specify a data revision code that tracks updates to the data. This also requires documentation of those updates (e.g., new calibration, timing error, etc.) to be recorded in the file header (see section 2.3.B). For this we specify a revision number counter “_R#” where the underscore is a required element to separate the fields (this is needed for certain file checking software). The revision number “#” must match the revision number specified in the Normal Comments section of the file header (see section 2.3.B).

The optional parameters “_L#” and “_V#” may be needed in some special cases. If the contents of the file pertain to a second or third aircraft launch on the indicated date, then a launch counter “_L#” (i.e. L2, L3, etc.) must appear after the “R” identifier but before a volume counter, if present (see below). Launch number one is implied when “_L#” is omitted from the file name. If a data file is one volume of a multi-volume dataset, then a volume counter “_V#” (i.e. V1, V2, V3, etc.), must appear after the “R” parameter (and the “L” parameter, if present) separated by an underscore from the rest of the identifier. The volume number (the “#” in “V#”) must match the volume number in the file header. When “_V#” is missing from the file name a one-volume dataset is implied.

The optional comments parameter is for additional information required by the PI (or Data Manager) to identify the file contents but that does not fit into the other fields of the file name. This should be used sparingly.

2.3. File format specification for ICARTT time-series data files

2.3.A. Structure

The ICARTT time series data file format is structured to mimic the Ames file format **File Format Index (FFI) = 1001**. The definition of FFI in the Ames format is as follows: *The File Format Index (FFI) is used to uniquely define the exchange file format. By reference to pre-defined format options, the value of the FFI determines the number of INDEPENDENT variables, whether the values of the INDEPENDENT and dependent variables are numeric or character string, the format of the file header, and the format of the data records.*

We recommend that, whenever possible, ICARTT time series data files conform to the file format **FFI = 1001**.

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

This indicates that there is one independent variable, usually start time, and that all other data depend on the independent variable. 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. The format can also be condensed. For example, if measurements are reported continuously at 1 second intervals or less, then stop time and mid-point time need not be included. 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 section 2.3.B). As pointed out earlier, 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.

2.3.B. File header information

For the ICARTT data format, additional information is required and included in the comments sections. The most general header is shown below as an example; more specialized headers are described as modifications to the general form. Delimiters to separate fields (items) are commas only. For delimiters to separate text within an item, use underscores. The order in which data appears in the header is listed below. Words appearing in bold text are expected to appear in the header followed by the relevant information. Relevant example headers are provided following this list.

- Number of lines in header, file format index (most files use 1001) - comma delimited.
- PI last name, first name/initial.
- Organization/affiliation of PI.
- Data source description (e.g., instrument name, platform name, model name, etc.).
- Mission name (usually the mission acronym).
- 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 set to 1, 1) - comma delimited.
- UTC date when data begin, UTC date of data reduction or revision - comma delimited (yyyy, mm, dd, yyyy, mm, dd).
- 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 and for 10 Hz data the value is 0.1. All intervals longer than 1 second must be reported as Start and Stop times, and the Data Interval value is set to 0. The Mid-point time is required when it is not at the average of Start and Stop times. For additional information see Section 2.5 below.).
- Description or name of independent variable (This is the name chosen for the start time. It always refers to the number of seconds UTC from the start of the day on which measurements began. It should be noted here that the independent variable should monotonically increase even when crossing over to a second day.).
- Number of variables (Integer value showing the number of dependent variables: the total number of columns of data is this value plus one.).
- Scale factors (1 for most cases, except where grossly inconvenient) - comma delimited.
- Missing data indicators (This is -9999 (or -99999, etc.) for any missing data condition, except for the main time (independent) variable which is never missing) - comma delimited.

- Variable names and units (Short variable name and units are required, and optional long descriptive name, in that order, and separated by commas. If the variable is unitless, enter the keyword "none" for its units. Each short variable name and units (and optional long name) are entered on one line. The short variable name must correspond exactly to the name used for that variable as a column header, i.e., the last header line prior to start of data.).
- 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's and -8888's within the data, respectively. The last line of this section should contain all the "short" variable names on one line. The key words in this section are written in **BOLD** below and must appear in this section of the header along with the relevant data listed after the colon. For key words where information is not needed or applicable, simply enter N/A.).

The scanning program looks for these key words (case insensitive).

PI_CONTACT_INFO: Phone number, mailing address, and 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: Data Manager -- 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 file names discussion.

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. Pre-pend 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.

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 is: 14 + (# of dependent variables, given in line 10) + (# lines of special comments) + (# lines of normal comments).

2.3.C. Examples

Below are four examples of (similar) time series data using different forms of header information. 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: HOX_DC8_20040712_R0.ict

```
36, 1001
Brune, William
Penn State University
ATHOS - OH and HO2 concentrations using cryo water mix ratio data for quenching
corrections
ICARTT_INTEX
1, 1
2004, 07, 12, 2005, 01, 12
0
Start.UTC, seconds
4
1, 1, 1, 1
-9999, -9999, -9999, -9999
Stop.UTC, seconds
Mid.UTC, seconds
OH_pptv, pptv
HO2_pptv, pptv
0
18
PI_CONTACT_INFO: Address: 503 Walker Building, University Park, PA 16802; email:
brune@essc.psu.edu;
PLATFORM: NASA DFRC DC8 - sampling underneath aircraft forward cargo bay location
LOCATION: Aircraft location data in nav_dc8_20040712_R0.ict file
ASSOCIATED_DATA: see ftp://ftp-air.larc.nasa.gov/pub-air/INTEXNA/
INSTRUMENT_INFO: OH/HO2 LIF
DATA_INFO: Units are pptv.
UNCERTAINTY: The absolute accuracy is conservatively estimated to be +/- 32% at two
sigma confidence
ULOD_FLAG: -7777
ULOD_VALUE: N/A
LLOD_FLAG: -8888
LLOD_VALUE: N/A
DM_CONTACT_INFO: Bob Leshner; Penn State University; blesher@psu.edu
PROJECT_INFO: INTEX Mission 26 June-14 August 2004; California, Illinois, and New
Hampshire
STIPULATIONS_ON_USE: Use of these data requires prior approval from William Brune
OTHER_COMMENTS: N/A
REVISION: R0
R0: Final Data
Start.UTC, Stop.UTC, Mid.UTC, OH_pptv, HO2_pptv
55526, 55545, 55535, 0.171, 9.791
55546, 55565, 55555, 0.180, 9.218
55566, 55585, 55575, 0.186, 9.767
55586, 55605, 55595, 0.176, 9.996
55606, 55625, 55615, 0.192, 9.513
55626, 55645, 55635, 0.185, 9.798
55646, 55665, 55655, 0.160, 9.834
```

Example 2. All required data columns are shown explicitly.

File name: NOx_RHBrown_20040830_R0.ict

```
41, 1001
Williams, Eric
Earth System Research 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, seconds, number_of_seconds_from_0000.UTC
9
1, 1, 1, 1, 1, 1, 1, 1, 1
-9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999
Stop.UTC, seconds
Mid.UTC, seconds
DLat, deg_N
DLon, deg_E
Elev, meters
NO_ppbv, ppbv
NO_1sig, ppbv
NO2_ppbv, ppbv
NO2_1sig, ppbv
0
18
PI_CONTACT_INFO: 325 Broadway, Boulder, CO 80305; 303-497-3226;
email:eric.j.williams@noaa.gov
PLATFORM: NOAA research vessel Ronald H. Brown
LOCATION: Latitude, longitude and elevation data are 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 are 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_ppv, 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
```

2.4. File format specification for ICARTT multi-dimensional data files

2.4.A. Structure

ICARTT multi-dimensional data file formats are designed based on Ames standard file formats FFI=2110 and FFI=2310; we recommend using these FFI's for exchange of most multi-dimensional data files. The FFI descriptor is:

FFI 2110; two real independent variables, one unbounded and one bounded, with its values recorded in the data records; primary variables are real; the first auxiliary variable is NX(m,1) (or, primary variables' ArrayDimension), all other auxiliary variables are real.

FFI 2310; two real independent variables, one unbounded and one bounded, with its number of constant increment values, base value, and increment defined in the auxiliary variable list; primary variables are real; auxiliary variables are real.

For more details on these file types, please see the following documents:

<http://www-air.larc.nasa.gov/missions/etc/Amend2110.htm>

<http://www-air.larc.nasa.gov/missions/etc/Amend2310.htm>

2.4.B. Examples

Below are two examples on types FFI 2110 and FFI 2310

Example: FFI 2110

File name: AR_DC8_20050203_R0.ict

```
54, 2110
PI LastName, First Name
Code 916, Goddard Space Flight Center, Greenbelt, MD 20771
AROTAL
PAVE Mission
1, 1
2005, 02, 03, 2006, 01, 18
1
Altitude[], meters, Altitude_array
UTC, XX.XXXX_hours_from_0_hours_on_flight_date
7          ;{Number of PRIMARY variables}
0.1, 0.0001, 0.1, 0.01, 0.0001, 0.1, 0.0001
-99999, -999999, -999999, -999999, -999999, -999999, -999999
TempK[], K, Temperature_array
Log10_Density[], part/cc, Log10_NumDensity_array
TempK_Err[], K, Temperature_error_array
```

```
AerKlet[], Klet, Aerosol_array
Log10_O3NumDensity[], part/cc, Log10_Ozone_NumDensity_array
O3_MR[], ppb, Ozone_mixing_ratio_array
Log10_O3NumDensity_Err[], part/cc, Log10_NumDensity_error_array
11      ;{Number of AUXILIARY variable}
1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0
-9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999
NumAlts, none, Number_of_altitudes_reported
Year, UT
Month, UT
Day, UT
AvgTime, xxx.x_minutes, Averaging_time_of_presented_data
Latitude, degrees
Longitude, degrees
PAlt, meters, pressure_altitude
GPSAlt, meters, GPS_altitude
SAT, K, Static_air_temperature
SZA, degrees
0
18
PI_CONTACT_INFO: Enter PI Address here
PLATFORM: NASA DC8
LOCATION: Lat, Lon, and Alt included in the data records
ASSOCIATED_DATA: N/A
INSTRUMENT_INFO:
DATA_INFO:
UNCERTAINTY: Contact PI
ULOD_FLAG: -7777
ULOD_VALUE: N/A;
LLOD_FLAG: -8888
LLOD_VALUE: N/A;
DM_CONTACT_INFO: Enter Data Manager Info here
PROJECT_INFO: PAVE MISSION: Jan-Feb 2005
STIPULATIONS_ON_USE: Use of these data should be done in consultation with the PI
OTHER_COMMENTS:
REVISION: R0;
R0: Version 2005-0: AROTAL T & O3 Rayleigh Retrievals. Further revisions may be
needed to fine-tune aerosol characterization.
UTC, NumAlts, Year, Month, Day, AvgTime, Latitude, Longitude, PAlt, GpsAlt, SAT, SZA,
Altitude[], TempK[], Log10_NumDensity[], TempK_Err[], AerKlet[], Log10_O3NumDensity[],
O3_MR[], Log10_O3NumDensity_Err[]
54000, 9, 2005, 2, 3, 0, 42.308, -70.582, 6910, 6979, 242.5, 65.5
    9154, -9999, -999999, -9999, -9999, 113178, 212, -999999
    9304, -9999, -999999, -9999, -9999, 123353, 2250, -999999
    9454, -9999, -999999, -9999, -9999, 123008, 2116, -999999
    9604, -9999, -999999, -9999, -9999, 120933, 1337, -999999
    9754, -9999, -999999, -9999, -9999, 119675, 1019, -999999
    9904, -9999, -999999, -9999, -9999, 122655, 2061, -999999
    10054, -9999, -999999, -9999, -9999, 124384, 3126, -999999
    10204, -9999, -999999, -9999, -9999, 124632, 3371, -999999
    10354, -9999, -999999, -9999, -9999, 121341, 1609, -999999
54001, 8, 2005, 02, 03, 0, 42.278, -70.613, 6978, 7043, 241.7, 65.5
    10118, 9999, -999999, -9999, -9999, 124458, 3205, -999999
    10268, -9999, -999999, -9999, -9999, 123160, 2421, -999999
    10418, -9999, -999999, -9999, -9999, 121221, 1582, -999999
    10568, -9999, -999999, -9999, -9999, 120950, 1523, -999999
    10718, -9999, -999999, -9999, -9999, 117339, 680, -999999
```



```
10868, -9999, -999999, -9999, -9999, 122751, 2423, -999999
11018, -9999, -999999, -9999, -9999, 124230, 3491, -999999
11168, -9999, -999999, -9999, -9999, 124039, 3424, -999999
```

{Note the use of scale factors in this example.}

Example 2310

File name: LIDARO3_WP3_20040830_R0.ict

```
46, 2310
Williams, Eric
NOAA/Earth System Research Laboratory
Ozone number density profile from WP3 aircraft LIDAR
ICARTT_ITCT
1, 1
2004, 08, 30, 2009, 09, 04
1
Geo_Alt, meters, Geometric_altitude_of_observation
UT_TIME, seconds, Elapsed_time_from_0_hours_on_day_given_by_date
1 ;{Number of PRIMARY variables}
1.0e9
-9999
O3_NumDensity[], molecules/cc, Ozone_NumDensity_Array
9 ;{Number of AUXILIARY variable}
1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0
-9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999
Num_Altitudes, none, number_of_altitudes_at_current_time mark
Geo_Alt_Begin, meters, geometric_altitude_at_which_data_begin
Alt_Increment, meters, altitude_increment_between_observations
Geo_Alt_Aircraft, meters, geometric_altitude_of_aircraft
UT_hour, hours
UT_min, minutes
UT_sec, seconds
Lon_aircraft, degrees_E
Lat_aircraft, degrees_N
0
18
PI_CONTACT_INFO: 325 Broadway, Boulder, CO 80305; 303-497-3226;
eric.j.williams@noaa.gov
PLATFORM: NOAA WP3
LOCATION: Lat, Lon, and Alt data included in the data records
ASSOCIATED_DATA: N/A
INSTRUMENT_INFO: Differential absorption LIDAR. See Williams et al., BigScience, 42,
p. 50-51, 2001
DATA_INFO: The units are number density (#/cc). The vertical averaging interval is
975 m at 1-7 km above the aircraft and 2025 m > 7 km above the aircraft.
Horizontal averaging interval: 60 km.
UNCERTAINTY: Contact PI
ULOD_FLAG: -7777
ULOD_VALUE: N/A
LLOD_FLAG: -8888
LLOD_VALUE: N/A
```

```
DM_CONTACT_INFO: N/A
PROJECT_INFO: ICARTT study; 1 July-15 August 2004
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.
UT_TIME, Num_Altitudes, Geo_Alt_Begin, Alt_Increment, Geo_Alt_Aircraft, UT_hour,
UT_min, UT_sec, Lon_aircraft, Lat_aircraft, O3_NumDensity[]
30335, 26, 12819, 75, 10389, 8, 25, 35, -133.24, -9.45
    1340, 1519, 1660, 1779, 1868, 1939, 1973, 1992, 1989, 1955, 1934, 1897, 1817,
    1721, 1619, 1514, 1434, 1343, 1258, 1203, 1140, 1088, 1037, 956, 892, 878
30336, 22, 12819, 75, 10383, 8, 26, 0, -133.22, -9.93
    1351, 1523, 1658, 1774, 1860, 1930, 1962, 1974, 1966, 1932, 1909, 1877, 1803,
    1706, 1600, 1493, 1407, 1310, -9999, -9999, 1094, 1045
```

{Note that this file uses a scale factor (1e9) for the number density data since it would be very cumbersome to add the exponential notation to every value. Also, this example was adapted from the NASA document and did not have uncertainty or flag values associated with the data.}

2.5. File formats for non-standard airborne data

Data acquired by sensors on satellites are not conveniently incorporated into the ICARTT format. The data format allows each data record to be identified with a single timestamp only if data are reported continuously with a constant time interval (e.g., 1 second). Otherwise, start and stop times must be reported, and a data interval of 0 is entered on line 8 of the file header. Satellite data are unique in that while they are recorded on a constant data interval, significant gaps in the data may exist. These gaps may be due to cloud interference, changes in viewing mode (e.g., nadir versus limb), or other considerations. Given the sheer volume of data and the file sizes associated with satellite observations, it is not sensible to populate these data gaps with missing data values. It is also unreasonable to report start and stop times since data are typically collected on short timescales (typically sub-second) such that integration time is not an issue. Instead, satellite data files report a **Data Interval** of -1 on line 8 of the file header. This signifies that each data record is identified by a single timestamp, but the actual timeline is discontinuous. The ICARTT format does not support a **Data Interval** of -1 for any measurements other than from satellite instruments.

In some cases, the standard ICARTT time-series format does not easily conform to certain non-standard data. The data management team should consider, on a case-by-case basis, to use standards common to the user community, contingent upon agreement by the mission Science Team. For example, many modeling data sets store data in NetCDF (Network Common Data Form) format, which is a de facto standard for that community. However, the multi-dimensional data format defined above can accommodate these data sets, and we leave this as an optional format. For some instruments (e.g., lidars), data are available as image files usually in standard formats such as GIF or JPEG. Not all software for reading and writing these formats allow additional text information (e.g., as a header) so the file names for these files must be defined to

include as much information as possible. If necessary, the data management team should work with these PIs to achieve a mutually acceptable solution.

2.6. File scanning software

A software package “FScan” has been developed for scanning data files and verifying if the files are in compliance with the ICARTT format standards. The scanning function does a thorough examination on the file to ensure compliance; the file is checked line-by-line, value-by-value, and in some cases letter-by-letter. A detailed report is generated displaying error messages along with line numbers and reasons, if any. The “FScan” offers both online and standalone versions (see URL below). Further details on FScan is given at:

<http://www-air.larc.nasa.gov/missions/etc/helpFscan.html>.

There are 2 versions available to scan ICARTT formatted files:

1. Web-based: <http://www-air.larc.nasa.gov/cgi-bin/fscan>
2. Standalone Version (Windows only): <http://www-air.larc.nasa.gov/missions/etc/wFscan.htm>

3 References

Normal References

[1] NASA Ames Format Specification for Data Exchange:
http://espoarchive.nasa.gov/archive/docs/formatspec_2_0.html

[2] NASA GTE Data Archive Format:
http://www-gte.larc.nasa.gov/trace/TP_APP-E.htm

[3] Official ICARTT Data Format Document:
<http://www-air.larc.nasa.gov/missions/etc/IcarttDataFormat.htm>

Informative references

[4] International Consortium for Atmospheric Research on Transport and Transformation (ICARTT) campaign:
<http://www.esrl.noaa.gov/csd/ICARTT/>

[5] ARCTAS Data Policy and Management Plan:
http://www-air.larc.nasa.gov/missions/arctas/docs/arctas_data_plan.pdf

[6] MILAGRO Data Policy and Management Plan:
http://www.eol.ucar.edu/projects/milagro/data/MILAGRO_DataPolicy.html

4 Authors' Address

Ali Aknan, ali.a.aknan@nasa.gov, NASA/LaRC, MS 927, Hampton, VA 23681

Gao Chen, gao.chen@nasa.gov, NASA/LaRC, MS 483, Hampton, VA 23681

James Crawford, james.h.crawford@nasa.gov, NASA/LaRC, MS 483, Hampton, VA 23681

Eric Williams, eric.j.williams@noaa.gov, NOAA/ESRL, 325 Broadway, Boulder, CO 80305

5 Appendix A

Glossary of acronyms

<u>Acronym</u>	<u>Description</u>
ARCTAS	Arctic Research of the Composition of the Troposphere from Aircraft and Satellites
EU	European Union
GPS	Global Positioning System
GTE	Global Tropospheric Experiment
Hz	Hertz
ICARTT	International Consortium for Atmospheric Research on Transport and Transformation
INTEX-B	Intercontinental Chemical Transport Experiment – Phase B
INTEX-NA	Intercontinental Chemical Transport Experiment - North America
ITCT	Intercontinental Transport and Chemical Transformation
ITOP	Intercontinental Transport of Ozone and Precursors
LIDAR	LIght Detection And Ranging
MILAGRO	Megacity Initiative: Local and Global Research Observations
N/A	Not Applicable
NASA	National Aeronautics and Space Administration
NEAQS	New England Air Quality Study
NOAA	National Oceanic and Atmospheric Administration

NSF	National Science Foundation
PI	Principal Investigator
POLARCAT	POLar study using Aircraft, Remote sensing, surface measurements and modelling of Climate, chemistry, Aerosols and Transport
QA/QC	Quality Assurance / Quality Control
UTC	Universal Time Coordinated