NASA Aerogeophysics ASCII File Format Convention

Version 1.0, September 25, 2013

Operation IceBridge produces a wide range of data products that vary greatly in volume and complexity. Many of the L1B and L2 data products have been originally provided in ASCII file formats without following a standardized format. This document attempts to define a convention for ASCII file formats that can accommodate a wide range of data product requirements without being overly restrictive. The proposed IceBridge ASCII file format is intended to accommodate the needs and requirements of aerogeophysical missions. However, it might also be a suitable alternative for other projects using the more complex ICARTT format which was specifically designed for atmospheric science missions. The aerogeophysics ASCII file format convention incorporates many elements of the existing ICARTT standard.

1) General File Format Specification and Structure

The aerogeophysics ASCII file format is primarily intended for low-volume time-series data with parameters being measured sequentially (and/or simultaneously) in time. It is also suitable for L1B and L2 two-dimensional (i.e., along-track) derived geophysical products that have been interpolated onto a common geographic location and/or time base The use of this format convention is not limited to any level of data products. Data must be stored using the American Standard Code for Information Interchange (ASCII) character-encoding scheme and organized as a matrix of rows and columns. The data section of the file is comprised of ASCII alphanumeric characters (including scientific notations). Delimiters between columns of data include commas, tabs, or spaces. If fixed-width data columns are desired, use multiple spaces between data columns as padding and for the purpose of visual clarity (alignment) of data. Rows are separated by the end-of-line (EOL) character(s) for text files specific to the operating system. All rows must have an associated geographic location and time tag as specified below. All data points (variables) in a row must have the same geographic location and/or time tag. The numeric representation of a variable is defined by the units in which it was measured or derived. All variables must be represented using units of measurement approved by the International System of Units (SI system), derived units (such as degree Celsius) or non-SI units accepted for use with SI (such as minute, hour, day).

The data section must be preceded by at least one or more header lines each starting with number sign (#) that contains documentation about the data stored in rows and metadata about the file.

There is no restriction on how many header lines, rows and columns a data file can have. However, the file size is not to exceed 1 GB due to the current maximum allowed file size of the NSIDC ECS system.

2) Location Information

All L1B and L2 data products need to have an associated geographic location including latitude, longitude, and elevation. All data points (variables) in a row must have the same geographic location and all rows must contain geographic location information. 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). Alternatively, longitudes can be provided in either the 0° – 360° range as well. Elevation is to be reported in meters. The type of elevation measurement being used must be explicitly identified in the header metadata since many types of elevation measurements are available and in common use (e.g., GPS elevation; radar elevation; pressure elevation, etc.). Precise geodetic information requires an underlying reference frame and reference ellipsoid. The reference frame, such as the International Terrestrial Reference Frame (ITRF) and its epoch (e.g., 2008) must be explicitly defined in the header section and in a separate document that accompanies the data file as well as the reference ellipsoid, such as WGS-84, GRS-80, or TOPEX. The metadata must also document information about the type of location information that was used to create latitude, longitude, and elevation information for each row. For example, some data products are created using real-time GPS feeds into a logging computer, while other products are created using post-processed DGPS or PPP solutions interpolated onto the time tags of the measurements and their variables. The metadata must also document the location of the GPS antenna on the aircraft and whether antenna positions have been used for the geolocation or if the locations in the data file reflect sensor/instrument positions.

3) Time Information

All dates and times should be in UTC. Use of GPS time is discouraged. The preferred representation of dates and times should follow the ISO 8601 standard (e.g., 'yyyymmddTHHMMSSZ'). Use of seconds of the week is also discouraged. All data points (variables) in a row must have the same time stamp and consecutive rows must have monotonically increasing time tags. L3 and higher level data products that have been spatially interpolated are excepted from this.

All rows must contain information including year (four digits), date and time. A date can be represented as either day of year (3 digits) or a combination of month (two digits) and day of month (two digits). Time can be represented as hours (two digits), minutes (two digits) and decimal seconds. Timestamps can also be reported as UTC decimal seconds from the start of the date on which measurements began. If UTC seconds past midnight are used the reported time should be monotonically increasing even when crossing over to a second day (i.e., >86400). If this is the case then the date should not roll over so that when the UTC seconds are added to the date information the correct time tag is produced.

The processing steps to derive certain higher-order geophysical products may not make it possible/practical to keep timestamps with the data product. In that case, the start and stop times reflecting the data collection window should be included in the header metadata.

The time base information must be explicitly defined in the header section and/or in a separate document that accompanies the data file. Information regarding the time base should include the source of the time stamps (e.g. real time kinematic GPS, logging computer time). If GPS times are used, then the location of the GPS antenna on the aircraft needs to be documented. If internal computer clocks have been used, the documentation should include whether these clocks have been synchronized to GPS time or not.

4) Missing Data

Missing data should be represented by negative numbers large enough to never be construed as actual data, such as -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.

The missing data values, LOD, etc. need to be documented in the header section of the data file.

Alternatively, NaN (not a number) can be used to indicate missing data values as well.

5) Header Information

The data section must be preceded by at least one or more header lines starting with the number sign (#) that contain documentation about the data and metadata about the file. There is no restriction on how many header lines a data file can have, however, the number of header lines for a particular data product should not change over the course of the project. At a minimum the header must contain at least one line that lists the unique names of the variables (columns) separated by one of the three delimiters (see section 1). If missing data values or any other flags described in section 4) are used, a separate header line must define the flags and values. In order to accommodate a wide range of complexities two end member scenarios (see below) are envisioned for information to be included in the header.

a) Data products with a small number of variables (columns) that are more or less directly acquired by an instrument and require little documentation on data processing should aim to include all metadata in the file header. An example of such a data product is the MCoRDS L2 ice thickness data which contains nine columns (variables):

```
# The MCoRDS L2 Ice Thickness data set contains measurements for Elevation, Surface, Bottom and Thickness.
# Format is 9 columns separated by commas. Columns are:
#
# LAT                  Latitude    Degrees North, Antenna position on the aircraft w.r.t WGS-84 and ITRF2008 from postprocessed GPS data.
# LON                  Longitude   Degrees East, Antenna position on the aircraft w.r.t WGS-84 and ITRF2008 from postprocessed GPS data.
# TIME                 UTC Time    Seconds of day. Note: When aligning with GPS time tagged data, account for leap seconds.
#                      Note: time tages can be obtained by using the YYYY MM DD fields from the FRAME and adding TIME.
# THICK     Ice Thickness: Bottom minus Surface. Constant dielectric of 3.15 (no firn) is assumed for converting propagation delay into range. -9999 indicates no
thickness available.   Meters
# ELEVATION            Elevation of GPS antenna referenced to WGS-84 Ellipsoid and ITRF2008 .            Meters
# FRAME     (YYYYMMDDSSFFF) Fixed length numeric field. YYYY = year, MM = month, DD = day, SS = segment FFF = frame.
# SURFACE   Range to Ice Surface. Actual surface height is Elevation minus this number.        Meters
# BOTTOM    Range to Ice Bottom. Actual ice bottom height is Elevation minus this number. Constant dielectric of 3.15 (no firn) is assumed for converting propagation
delay into range. -9999 indicates no thickness available.  Meters
# QUALITY   1: High confidence pick
#               2: Medium confidence pick
#               3: Low confidence pick
#
# LAT,LON,TIME,THICK,ELEVATION,FRAME,SURFACE,BOTTOM,QUALITY
75.767666,-55.039845,42410.9208,1310.03,4046.8340,2012050804001,2318.54,3628.57,1
75.767737,-55.039383,42411.1282,1312.26,4046.4525,2012050804001,2318.54,3630.81,1
75.767816,-55.038861,42411.3357,1314.49,4046.0219,2012050804001,2318.54,3633.04,1
75.767957,-55.037936,42411.5432,1316.73,4045.2594,2012050804001,2318.54,3635.27,1
75.768098,-55.037004,42411.7507,1318.96,4044.4917,2012050804001,2318.54,3637.51,1
2012,102,39958.19,66.988627,309.031383,2966.88,-17.21,-1.53
```

b) Data products with a large number of columns and a complex processing flow that merges multiple L1B and L2 data products to derive new higher order data products including uncertainties should aim to provide as much information as possible in the header. It is not practical to include specifics about data processing etc in a header, because this information often requires 30-40 pages of documentation. The sea ice thickness and snow depth product (IDCSI2) is an example for such a complex product with a large number of variables:

```
#lat,lon,thickness,thickness_unc,mean_fb,ATM_fb,fb_unc,snow_depth,snow_depth_unc,n_atm,pcnt_ow,pcnt_thin_ice,pcnt_grey_ice,corr_elev,elev,date,elapsed,atmos_corr,geoi
d_corr,ellip_corr,tidal_corr,ocean_tide_corr_part,load_tide_corr_part,earth_tide_corr_part,ssh,n_ssh,ssh_sd,ssh_diff,ssh_elapsed,ssh_tp_dist,surface_roughness,ATM_fil
e_name,Tx,Rx,KT19_surf,KT19_int,low_en_corr,sa_int_elev,si_int_elev,my_ice_flag,empty1,empty2,empty3,empty4,empty5,empty6,empty7,empty8,empty9,empty10
    81.410301,   263.314270,      4.0937,      1.3230,      0.8235,      0.8235,      0.1169,      0.5640,      0.0570,  110,      0.000000,      0.000000,      0.000000,
0.8913,     10.0418,20110316, 45923.804687500,     -0.1634,      9.7704,      0.7134,      0.0700,     -0.0500,      0.0036,      0.1164,      0.0514,      79,      0.0229,
0.0030, 46627.531250000,  99507.8594,      0.418452, 20110316_124407.atm4cT3.qi,  1465.0,  1828.8,-99999.00,-99999.00,-99999.0000,-99999.0000,-99999.0000,      1.00,-
99999,-99999.00000,-99999.00000,-99999.00000,-99999.00000,-99999.00000,-99999.00000,-99999.00000,-99999.00000,-99999.00000,
```