Airborne Data Management Plan for DC3 Field Study

Official version, January 2, 2012

Introduction

This document is intended to provide a framework through which the Deep Convective Clouds and Chemistry (DC3) airborne data are effectively archived, managed, and shared. This document also addresses the transfer of data for archival in the Atmospheric Science Data Center at NASA Langley Research Center (LaRC ASDC). These data are also copied to the NCAR/EOL DC3 Data Archive Center (DDAC).

The DC3 data will undergo three stages within the project-lifecycle: field data, preliminary data, and final data. The field data are generated during the field deployment and are primarily used to measure progress in achieving the science goals. The preliminary data incorporate post-deployment instrument calibration and characterization, data synchronization, and the QA/QC (quality assurance / quality control) process. The data processing requires multiple PIs' data to finalize the QA/QC process (a step called integrated data processing and analysis) that may reveal issues requiring reevaluation of preliminary data products. The final data are intended to be publication quality and open to the public. As required by NASA's data policy, the final data will also be transferred to the LaRC ASDC. The overarching goal of the DC3 data management plan is to generate high quality science data, deliver data products in a timely manner, stimulate the interest of the scientific community, and ultimately help achieve the overall science objectives.

Data Repository

A data repository is established for the DC3 airborne observations at http://www-air.larc.nasa.gov. The Airborne Science Data for Atmospheric Composition (ASD-AC) group at NASA Langley Research Center is responsible for maintaining the data archive. As the project progresses, the airborne data archive will sequentially host: field data, preliminary data, and final data. The field data will be expunged after the preliminary data due date and preliminary data will be removed after the final data is due. The data submission schedule is given in the "Data Submission Timeline" section. The airborne data archive will host data from the NSF GV, NASA DC-8, and DLR Falcon-20. Ground-based observational data (radar, lightning mapping array data, and radiosondes) are archived at the NCAR EOL repository. The NCAR/EOL data archive will also hold operational weather products, satellite data products, meteorological forecasts, and will receive a copy of the airborne data archive.

The airborne field data and preliminary data will be open only to DC3 science teams and their collaborators. Access will be protected by a common username and password which will be required to download data from the archive. The final airborne data archive will be transferred to LaRC ASDC and NCAR/EOL DDAC. Data revisions will be tracked through data file revision numbers as required by ICARTT file naming convention.

Data Archive Structure and DataID: The DC3 aircraft data archive will be constructed with a three-tier directory structure. The top, root level tier identifies the mission name: 'DC3'. The second level tier is based on the platforms (for example, 'NASA DC8' or 'NSF GV') on which data will be collected and the third level tier is derived from PI names in each platform, using the naming convention: LASTNAME.FIRSTNAME.

Under each PI's directory, the data files are organized by the PI based on the type of the measurements or instruments. The primary discriminator for all data files in the PI's directory is implemented by a "dataID", which is assigned by PIs prior to field deployment. ***Note that all PIs are required to register their "dataIDs" prior to their data submission, regardless whether ICARTT or HDF files are used.*** Otherwise, the system will not recognize their files as valid data inputs. The "dataID" is an identifier of the data source, which is typically an acronym describing the measurement group, measured species, instruments or model, etc. The "dataID" is part of the ICARTT filename structure (see Appendix A).

As an example, in past studies, "DLH" was used as a "dataID" for diode laser hygrometer measurements of water vapor data; "LARGE-APSsd" was used to denote the LARGE group's measurement of aerosol size distribution using the APS measurement; and "STEM" was used for the STEM (Sulfur Transport and dEposition Model) model results. For the DC3 campaign, all "dataIDs" will be prefixed with "DC3-".

Data Submission Timeline

The data submission timeline is designed to facilitate collaborative research for achieving the overall mission science objectives. The submission deadlines given here follow the DC3 Data Policy that is found on the DC3 web page: http://www.eol.ucar.edu/dc3/

Mission Study Phases	Data Type	Submission Deadline
Field Deployment	Field data	24 hour after each flight
Post-Deployment	Preliminary Data	January 20, 2013
Public	Final Data	July 1, 2013

During the Field Phase, the DC3 instrument principal investigators are required to submit their data to the data repository within 24 hours of the flight. Exemption may be granted by the project leadership for certain measurements which require additional data-processing time or when special circumstances occur, e.g., back to back-flights. The timely submission of field data is required to assess progress toward mission science goals and to plan subsequent flights. All field and preliminary data will be deleted when the preliminary data are delivered (January 20, 2013). The preliminary data will be removed after July 1, 2013.

The airborne preliminary data, due ~6 months after the field deployment, is primarily used for the integrated data processing and analysis, which serves as an important step toward finalizing the observational data. The airborne final data will be made publicly available on July 1, 2013 through the data repository and also transferred to the archives at LaRC ASDC and NCAR/EOL DDAC.

Data Format Requirements

DC3 airborne data format requirements are intended to facilitate seamless data exchange among the science team members and partners and to meet the standards for long-term data preservation. The airborne observational data products from in-situ measurements are required to conform to the International Consortium for Atmospheric Research on Transport and Transformation (ICARTT) data format standards. The ICARTT format is now one of the NASA Earth Science Division's approved data system standards (ESDS-RFC-019). A detailed description of the data format protocol can be found at http://www-air.larc.nasa.gov/missions/etc/IcarttDataFormat.htm. As required by the ICARTT format protocol, all DC3 airborne observational data must be reported with universal time (UT) for the time record. In addition, the DC3 project has a specific file naming convention to identify the airborne campaign, i.e., the file names will be prefixed with "DC3" (Detailed information is given in Appendix A). These additional requirements are needed for the LaRC ASDC archive and to promote the data usability.

All incoming data files will be electronically scanned to ensure compliance with the ICARTT format requirements. The scanning software will provide error messages if deviation from the ICARTT format is detected. Additional assistance will be made available through ASD-AC staff to the science team to trouble-shoot issues in generating ICARTT files.

The DC3 airborne remote sensing data products may opt to use ICARTT or HDF format. The HDF files must comply with one of the following format standards: HDF 5, HDF-EOS 5 or the HDF and HDF-EOS Profile heritage standard. More information can be found at: http://earthdata.nasa.gov/our-community/esdswg/standards-process-spg/rfc/esds-rfc-007; http://earthdata.nasa.gov/our-community/esdswg/standards-process-spg/rfc/ese-rfc-008; http://earthdata.nasa.gov/our-community/esdswg/standards-process-spg/docindexfolder/heritage/hdf-and-hdf-eos-profile. This reflects the fact that ICARTT cannot effectively handle arrays with more than 3 dimensions. To ensure data access to all, links to HDF Group/HDFView will be provided on the data archive website. As no specific metadata requirements are built into the HDF File Format protocols, DC3 PIs are required to provide the metadata equivalent to the ICARTT format metadata specifications, given in Appendix B. Like the ICARTT files, the HDF files will follow the naming convention given Appendix A. The incoming HDF files will be checked for the naming structure before being placed in the archive. UT should also be used in HDF files for reporting time of the observations.

Specific DC3 Data Reporting Requirements

In-situ measurement synchronization:

To ensure an accurate description of atmospheric phenomena, the in-situ measurement data products from the same aircraft platform are required to synchronize by using a common fast measurement. This synchronization process is considered as a correction for the difference in instrument time response and inlet delays. The science team has decided the reference time standards for each aircraft as given below:

Aircraft Platform	Reference Time Standard
NASA DC-8	DLH (Diode Laser Hygrometer)
NSF GV	Primary: VCSEL (vertical cavity surface emitting laser) Hygrometer Secondary: O3 (chemiluminescence O3)

Variable Naming standards:

As required by the ICARTT format protocol, each data variable shall have a short-name, unit, and an optional long-name, which is more descriptive. The DC3 airborne Science Team has adopted a consistent naming convention for the short variable names and uses long variable names along with the short names to enhance the data usability for a broad range of the scientific communities. The long variable names are intended to be more descriptive of the data reported and ideally should be consistent with the CF standard names (http://cf-pcmdi.llnl.gov/documents/cf-standard-names/standard-name-table/18/cf-standard-name-table.html). To streamline the variable naming process, a spreadsheet has been distributed to the airborne science team as the recommended variable short names, units, and long names for in-situ trace gas measurements. Similarly, in-situ aerosol measurement variable names and units have been provided in another spreadsheet. This information is posted on the airborne data archive website at NASA/LaRC. The DC3 airborne instrument PIs should choose the variable names and units from these spreadsheets. A suffix should be added to the short names If more than one measurement of the same species/parameter is on-board the same aircraft platform. In this case, the suffix will be in the form of "_" plus the Instrument/Group Acronym given by the PIs. For example, the DC-8 NO2 measurements may be named as NO2_LIF and NO2_CLD for data from laser induced fluorescence and chemiluminescence instruments, respectively.

Standards for in-situ trace gas data units:

There will be a few hundred trace gas measurements in the DC3 airborne field study. Many of these will be made on multiple platforms and by different instruments. It is recommended that consistent units be used in reporting the measurements of the same trace gas species. This is beneficial to the collaboration between the science teams and to users at large. The DC3 airborne data manager will work with the measurement PI groups to make recommendations. Recommended units for variables will be posted on the data archive website.

Standards for in-situ aerosol data units:

The DC3 airborne aerosol measurement PIs have reached consensus to report all data products under standard temperature and pressure (STP) conditions. The STP condition is defined as 1013 mb and 273.15 K. A conversion factor to ambient condition should be included in the preliminary and final data files as a data column. This requirement is to enhance the aerosol data reporting uniformity for DC3, which will help the data usability by broad science communities.

It has also been agreed upon that common units will be used for the same type of measurements. For example, particle number concentration will be reported in **cm^{-3}**; size distribution data in **dN/dlogDp** and in **cm^{-3}**; particle scattering and absorption coefficients in **Mm^{-1}**; and chemical composition data in **microgram std m^{-3}**, excepting black carbon which will be reported in **nanogram std m^{-3}**.

Science Data Guidelines

In order to ensure that data are used and acknowledged fairly and properly, all DC3 participants are required to follow the data policy on the DC3 web page. http://www.eol.ucar.edu/dc3/

Research Data Products

Combining all measurements from one platform or site on a common time base makes the files much easier to use for both data processing and interpretive analysis. Such merges are valuable for field data, preliminary data, and final data.

The ASD-AC team plans to create merges of the ICARTT-format files that are in the DC3 airborne data archive. The merge files will be made available at the data repositories. The merge files will be updated throughout the project lifecycle as the individual instrument PI data files are revised.

Airborne Data Manager

The DC3 Airborne Data Manager will monitor the airborne data submission status in accord with the data submission timeline. The airborne data manager will also coordinate the efforts to support implementation of ICARTT format and the production of the data merge files.

Airborne Data Manager Contact Information:

Gao Chen, NASA Langley Research Center, gao.chen@nasa.gov, 757-864-2290.

Appendix A. DC3 data file naming convention:

dataID_locationID_YYYYMMDD_R#.extension

The only allowed characters are: a-z A-Z 0-9_.- (that is, upper case and lower case alphanumeric, underscore, period, and hyphen). Fields are described as follows:

dataID: an identifier of measured parameter/species, instrument, or model (e.g., O3; NxOy; and PTRMS). For DC3 data files, the PIs are required to use "DC3-" as prefixes for their DataIDs, i.e., DC3-O3.

locationID: an identifier of airborne platform or ground station, e.g., GV, DC8. Specific locationIDs for each deployment will be provided on the data website.

YYYY: four-digit year

MM: two-digit month

DD: two-digit day (for flight data, the date corresponds to the UT date at take off)

R#: data revision number. For field data, revision number will start from letter "A", e.g., RA, RB, … etc. Numerical values will be used for the preliminary and final data, e.g., R1, R2, R3 … etc.

Extension: "ict" for ICARTT files, "h4" for HDF 4 files and "h5" for HDF 5 files.

For example, the filename for the DC-8 Diode Laser Spectrometer H_2O measurement made on June, 1, 2012 flight may be: DC3-DLH-H2O_DC8_20120601_RA.ict (for field data) or DC3-DLH-H2O_DC8_20120601_R1.ict (for final data)

Appendix B. Summary of ICARTT format metadata requirements (also required for HDF 5 files):

Platform and associated location data: Geographic location and altitude will be embedded as part of the data file or provided via a link to the archival location of the aircraft navigational data.

Data Source Contact Information: phone number, mailing information, and e-mail address shall be given for the measurement Co-I and one alternate contact.

Data Information: Clear definition of measured quantities will be given in plain English, avoiding the use of undefined acronyms, along with reporting units and limitation of data use if applicable.

Measurement Description: A simple description of the measurement technique with reference to readme file and relevant journal publication.

Measurement Uncertainty: Overall uncertainty is given as a minimum. Ideally, precision and accuracy is provided explicitly. The confidence level associated with the reported uncertainties is also specified for the reported uncertainties if it is applicable. The measurement uncertainty can be reported as constants for entire flights or as separate variables. Measurement uncertainty is required by the ICARTT data file format.

Data Quality Flags: definition of flag codes for missing data (not reported due to instrument malfunction or calibration) and detection limits.

Data Revision Comments: Provide sufficient discussion about the rationale for data revision. The discussions should focus on highlighting issues, solutions, assumptions, and impact.