Data Format and Metadata for NASA Airborne Observations: A Brief Overview

Gao Chen (Technical Lead) and Brandi Quam (Lead)
NASA Airborne Data Management Group

ET-WDC meeting, October 3, 2017

Airborne Data Formats

- NASA Airborne field studies have primarily used ICARTT, HDF, and netCDF formats
- ICARTT
  - Suitable for in-situ measurements
  - Easy to use for instrument scientist
  - File format integrity scanner available
  - Metadata not expandable and not machine comprehensible
  - File Header metadata requirements are somewhat vague
  - Inadequate to handle multi-dimensional data
- HDF and netCDF
  - Popular with modelers
  - Capable of handling all airborne observational data
  - Metadata expandable
  - Lack of standardized way to implement airborne metadata for data discovery, data preservation, and data use
  - Unfamiliar to most of the in-situ measurement scientists

ICARTT Format V1.0

- First developed in 2004 for the ICARTT field study by a group with extensive expertise in data use and data collection
- Consists of metadata file header and a data section, intended to be a simple and self-describing, ASCII file format
- The metadata attributes were designed for adequately describing in-situ measurements
- Had become widely accepted by airborne communities with primary focus on atmospheric composition measurements
- ICARTT V1.0 was adopted in 2010 as a standard format for airborne data, through ESDS Standard process working group

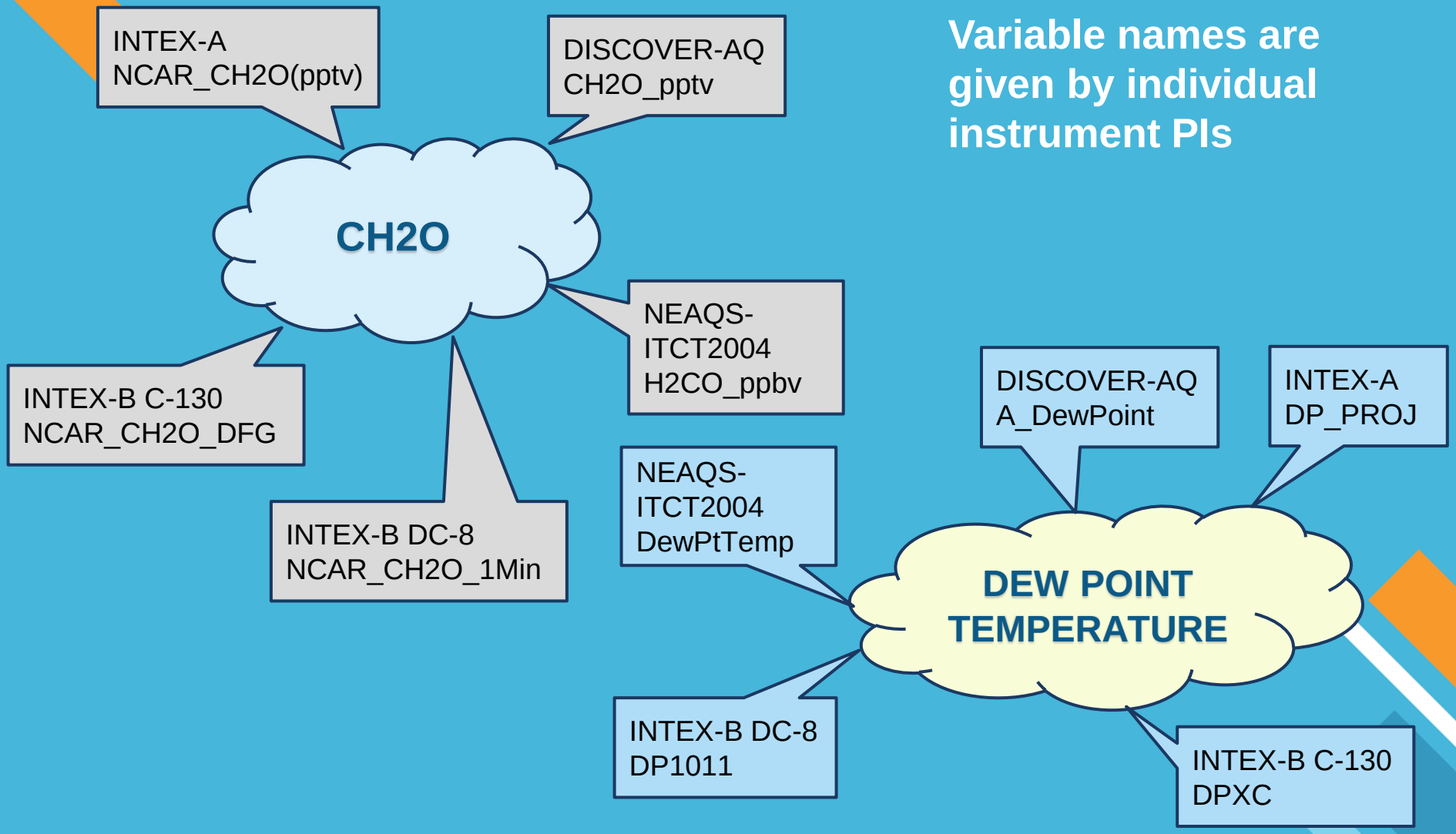
ICARTT Format V2.0

The ICARTT format was revised by the ESDS ICARTT Format Refresh working group to improve data usability and interoperability. The issues were identified by the users and the solutions were thoroughly discussed among the modelers and instrument scientists. The ICARTT V2.0 was formally approved in 2017.

- Require all variables to have variable standard name tags
- Standardized time variable names, i.e., Time_Start, Time_Stop, and Time_Mid for start, stop, and mid time of an integration interval
- Better structured file header metadata by introducing required keywords, order of the keywords, keyword/value pairs, and formatting for the keywords.

https://earthdata.nasa.gov/standards/icartt-file-format

Variable Standard Name: Current Status



Variable names are given by individual instrument PIs

Variable Standard Names: Ongoing Discussion

- **The Needs:** search and/or use data from different airborne field studies
- **The Role:**
  - ✓ Require all instrument PIs to use standard names
    - Pros: promote interoperability
    - Cons: Some analysis software (e.g., IGOR, IDL, and MatLab) can not handle multiple identical variable names; difficult to maintain individual data identities from multiple instruments when using ASCII format
  - ✓ Use standard names as tags
    - Pros: easy to maintain individual data set identities and work with all analysis software
    - Cons: variable name remains to be inconsistent for the same parameter
- **The approach to make a list of standard names:**
  - ✓ Discussions among data users, data providers, and data system developers to develop a list of ***community-acceptable*** standard names
  - ✓ A NASA Earth Science Data and Information System (ESDIS) working group has started to review the current available community variable name standards and to adopt/improve these standards, which will become the NASA standards that are linked to the WMO standards

Variable Standard Names: ICARTT V2.0 Implementation

ICARTT Variable Description Requirement:

VarShortName, unit, VarStandardName, VarLongName

- Variable Short Names
  - ✓ Selected by instrument PIs
  - ✓ Make recommendations: standard name with group or instrument identification?

    e.g., O3_NOAA_Ryerson, HNO3_CIMS_Huey
- Variable Standard Names
  - ✓ Simple and descriptive: Variable description + Medium
  - ✓ Inclusive – one standard name associate with multiple PI variables?
  - ✓ Lumped measurement variables?
- Variable Long Names
  - ✓ Need structure and content requirements

Metadata: Current NASA Standards

- NASA DAACs generally use Common Metadata Repository/Unified Metadata Model – Collection (i.e., directory) and Granule (e.g., file) standards (CMR/UMM-C and CMR/UMM-G), originally designed for satellite observations
- Each DAAC has its own interpretation to fit its ingest and distribution tools
- CMR/UMM metadata are primarily used for data discovery, distribution, and archival

CMR/UMM-C	P	D	U
Metadata Info	√	√	
Data Identification	√	√	
Descriptive Keywords		√	
Distribution Info		√	
Temporal/Spatial Info	√	√	
Acquisition Info	√	√	

CMR/UMM-G	P	D	U
Metadata Info	√	√	
Data Identification	√	√	
Temporal/Spatial Info	√	√	
Coverage Info	√	√	
Distribution Info		√	
Data Quality Info	√		√
Acquisition Description	√	√	

P = Preservation, D = Discovery, U = Use

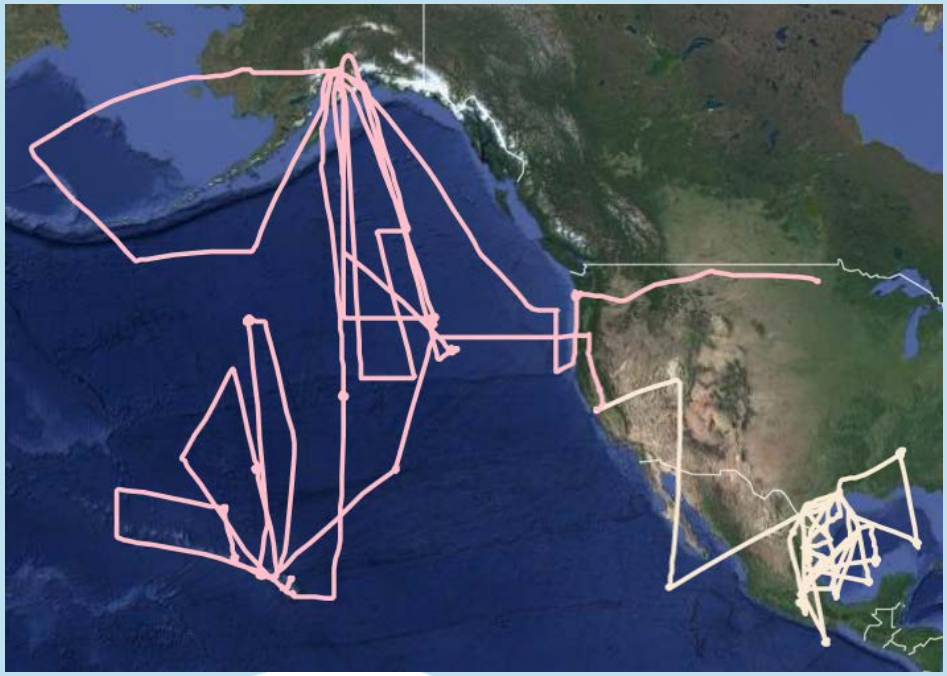
https://earthdata.nasa.gov/about/science-system-description/eosdis-components/common-metadata-repository/unified-metadata-model-umm

Proposed Discovery Metadata

- Spatial/Temporal Sampling Extent
- Mission and platforms
- Variable/Parameters
- Instruments
- Instrument PIs
- PI, platform, Instrument, and Variable relationship
- PI variable names and associated standard names
- Others?

Metadata Challenges: Sampling Region Representation

NASA INTEX-B DC-8 Flight Tracks



- The actual flight tracks are influenced by uncontrollable factors
- Data spatial search and/or subsetting should be based on actual flight tracks
- A current compromise to fit CMR/UMM requirements is to use multiple boxes to represent the general sampling region
- 3-D airborne sampling locations for airborne in-situ measurements are reported as data, not metadata

Proposed Data Use Metadata

- Variable definition
- Variable unit
- Variable type: scalar, vector, array, flags …
- Data Flags for missing data, LODs, …
- Measurement uncertainty reporting
- Measurement type specific metadata
  - ❑ in-situ trace gas metadata:
    - ✓ Lumped vs. specific measurement: NO vs. NOy
    - ✓ Type of measure: mass concentration, number concentration, mixing ratio
    - ✓ Reporting conditions: dry mixing ratio, STP…
  - ❑ in-situ aerosol metadata:
    - ✓ Inlet and instrument size cut
    - ✓ Size determination: optical, mobility, aerodynamic, …
    - ✓ Size bin definition
    - ✓ Reporting conditions: dry vs. ambient humidity; STP vs. ambient temperature and pressure
  - ❑ Active and passive remote sensing data ???

Metadata Challenges: Variable Description

- For many in-situ observed variables, additional information are needed to properly use the data. For example:
  - A CO2 measurement needs to be specified if the data are reported as dry mixing ratio or ambient mixing ratio
  - Aerosol mass concentrations need to be indicated if the data are reported under ambient or STP conditions
  - Aerosol size measurements need to be defined in terms of the type of size determination, e.g., optical, mobility, or aerodynamic diameter
- Based on discussions between modelers, instrument experts, and data scientists, additional metadata attributes were proposed for the new UMM-Variable standards:
  - Measurement conditions: the conditions under which the measurements were made and cannot be altered through data processing.
  - Reporting conditions: the conditions under which the variable is reported, e.g., STP, dry mixing ratio
  - Variable standard name tag

Proposed General Instrument Metadata for Preservation

- Instrument Name
- Airborne Platform
- PI Contact Information and Affiliation
- Measurement (variable list)
- Detection Principle
- Manufacture/Developer
- In-situ instrument calibration standard/methodology
- Remote sensing instrument???
- Publication/Description
- Airborne Configuration (e.g., Inlet)
- Specific Operational Note (e.g., instrument status for individual flight)
- Data Processing Note/Description (when applicable)???
- Data Revision Record(s)