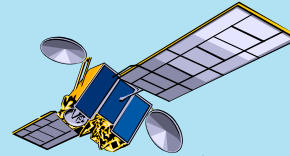


AIRBORNE DATA MANAGEMENT FOR NASA: Challenges and User-driven Strategies

Gao Chen and Brandi Quam
NASA Airborne Data Management Group

Jeff Walter and John Kusterer
Atmospheric Science Data Center, NASA Langley Research Center

An integrated approach involving observations and models promotes the understanding of atmospheric processes, which helps to mitigate air quality issues



- Small number of species/parameters
- Broad spatial coverage
- Long-term measurements with limited sampling times

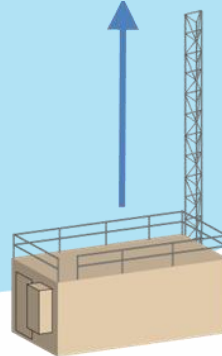


- Large number of measurements: trace gases and aerosol properties
- Limited spatial and temporal coverage with high resolutions

Model Assessment
Data Assimilation
Air quality Study
Transport Pathway Study
Radiative Forcing Study



- Global, regional, and process models for current and future scenarios



- Large number of measurements
- Continuous temporal coverage
- Limited spatial coverage

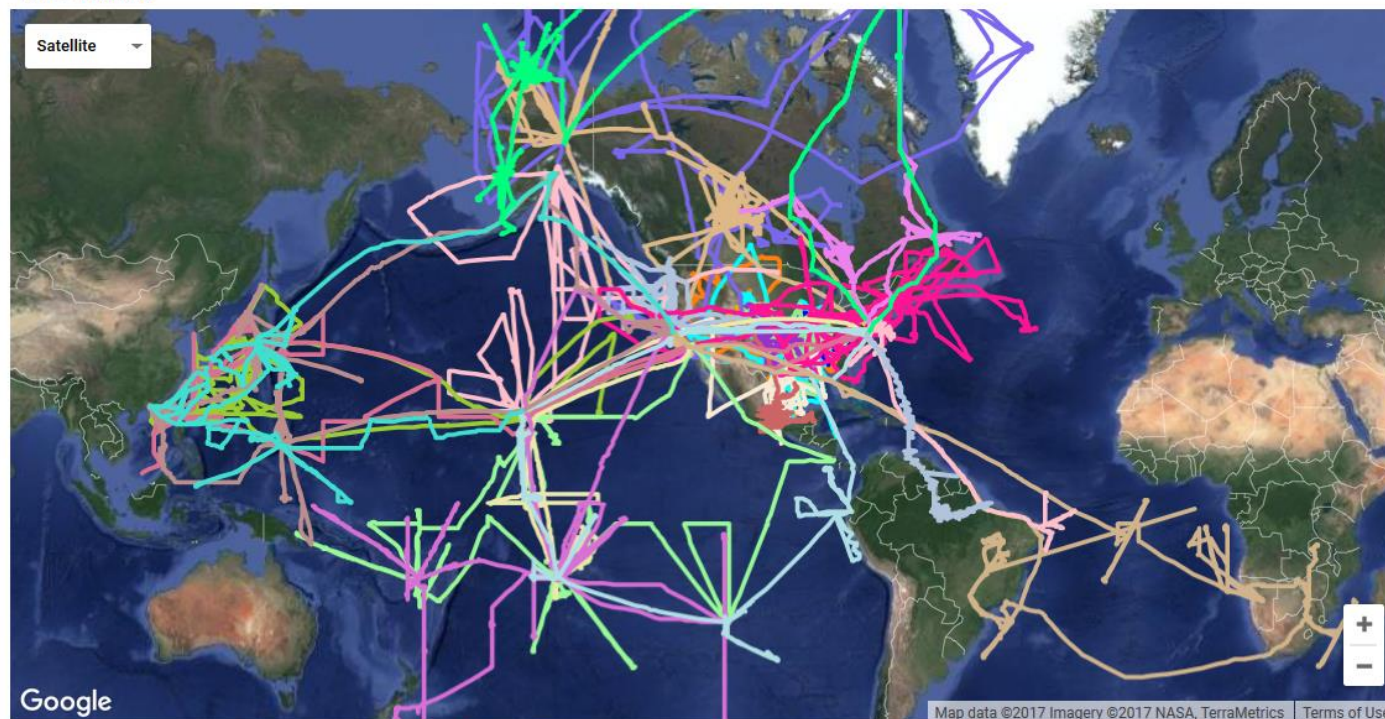
NASA Tropospheric Chemistry Campaigns -- Merged Data Sets

Each merged data set includes all or most measured species for a particular mission and applies a common time base to all of the data. For example, one merged set may index all of the data to a 1 second, 10 second, 1 minute or 5 minute time increment. Some merged sets are indexed by the time increment of a particular measured species, such as whole air samples or other measurements with extended integration times.

[- Hide Google Map](#)

Legend

discoveraq_CO_p3b
frappe_C130
discoveraq_TX_p3b
discoveraq_CA_p3b
discoveraq_MD_p3b
seac4rs_dc8
seac4rs_er2
dc3_dc8
dc3_gv
dc3_falcon
arctas_dc8
arctas_p3b
intexb_dc8_mx_tx
intexb_c130_mx
intexb_dc8_hi_ak
intexb_c130_wa
intexna_dc8
tracep_dc8
tracep_p3b
pemtropics_b_dc8
pemtropics_b_p3b
pemtropics_a_dc8
pemtropics_a_p3b
pemwest_b_dc8
tracea_dc8
pemwest_a_dc8
able_3b_electra
cite_3_electra
able_3a_electra
able_2b_electra



[- Shrink Map](#)

[Clear Map](#)

[Refresh Map](#)

[Cursor](#)

Lat: 74.56362

Lon: 89.67773

Select mission:

DISCOVER-AQ
2014 - 2011

SEAC4RS
2013

DC3
2012

ARCTAS
2008

INTEX-B
2006

INTEX-NA
2004

TRACE-P
2001

PEM-Tropics-B
1999

PEM-Tropics-A
1996

PEM-West-B
1994

TRACE-A
1992

PEM-West-A
1991

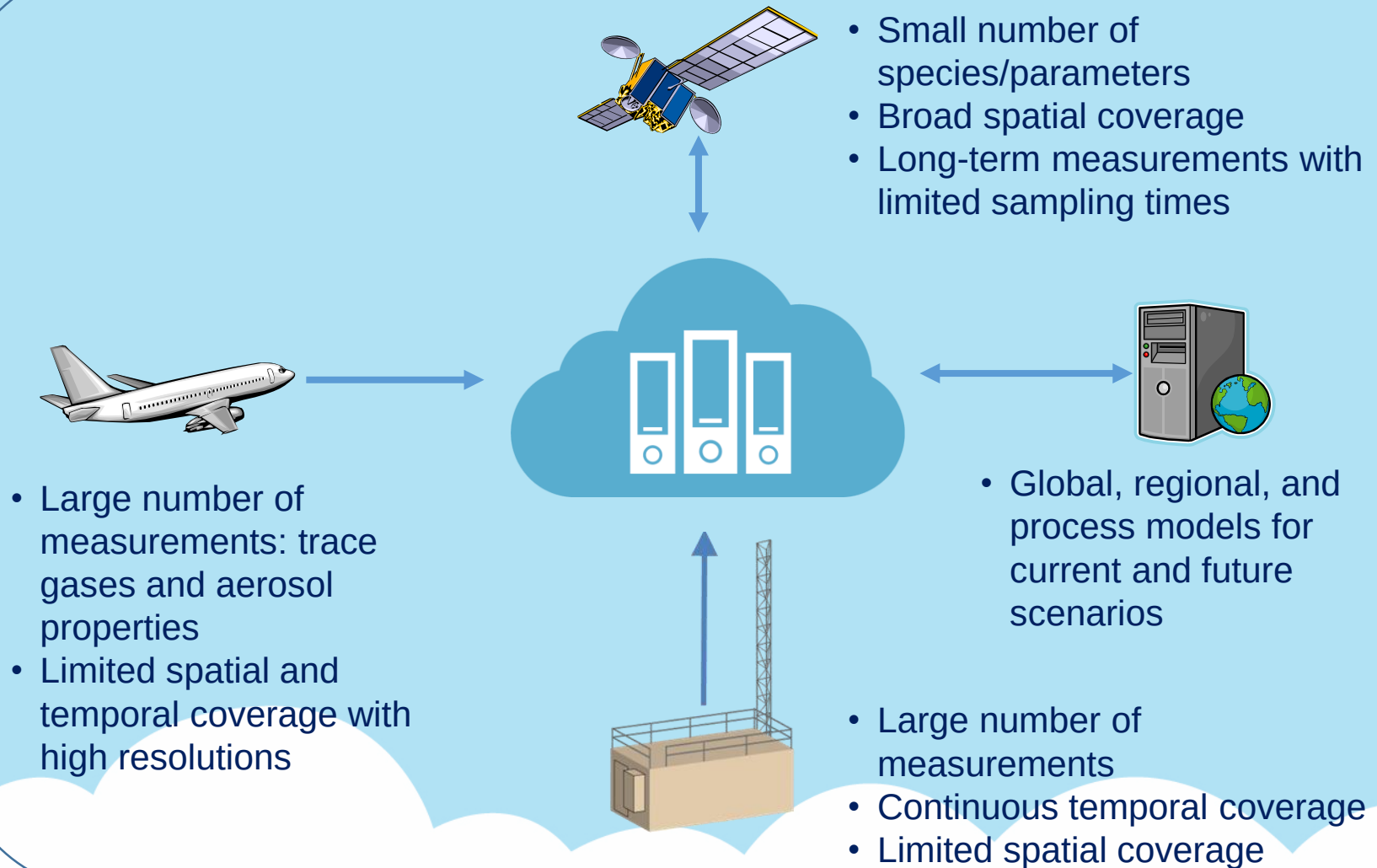
ABLE-3B
1990

CITE-3
1989

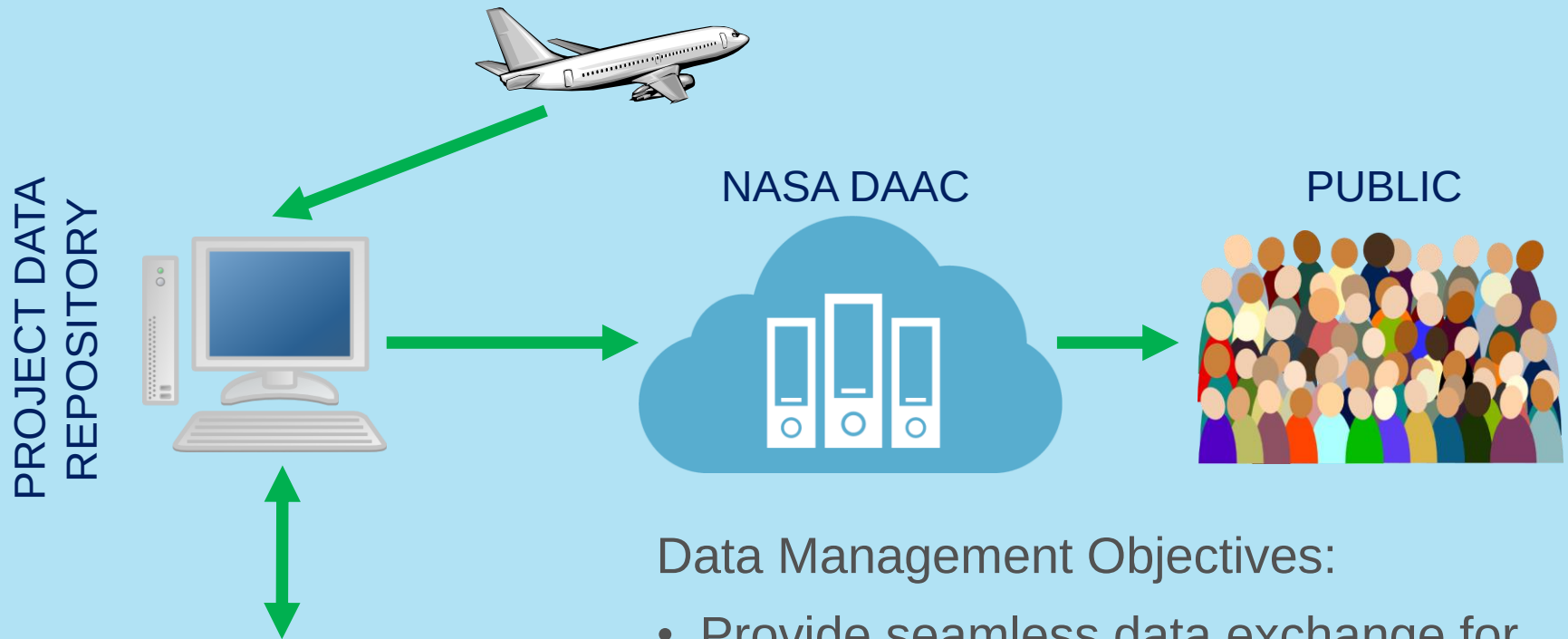
ABLE-3A
1988

ABLE-2B
1987

A data system is needed to integrate models and observations



NASA Airborne Data System - Idealized



SCIENCE TEAM

Data Management Objectives:

- Provide seamless data exchange for science team to facilitate data processing
- Provide data and metadata to Distributed Active Archive Center (DAAC) for long-term preservation and distribution

Project Data Repository



NASA Airborne Data System: Project Data Repository

A project data repository serves primarily for science team data exchange. This reflects that instrument data processing is co-dependent, i.e., data processing for one instrument may need data from another independent instruments

Field Data

Preliminary Data

Final Data

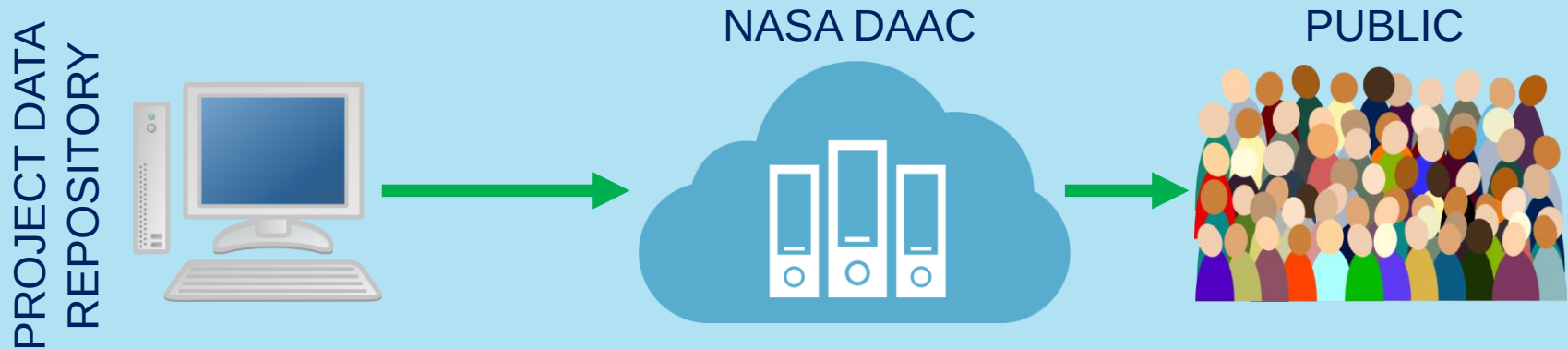
- **Field data:** quick-look data for flight planning and instrument operation checking
- **Preliminary data:** adjusted by post-deployment instrument calibration and/or characterization, initial QA/QC, and possible measurement timing synchronization between instruments
- **Final data:** publication quality, measurement timing synchronization between instruments



Science Team

Final data shall be immediately transferred to a DAAC and released to public

NASA Airborne Data System: Data Transfer to DAAC



- All data products from NASA airborne field studies are required to be transferred to an assigned DAAC for long-term preservation and public distribution purposes
 - Submit data management plan before deployment
 - Transfer the final data and the companion metadata files
- All NASA airborne data products are open public with no data use restrictions

Complexities with Airborne Data (I)

- from a data system's perspective

- A large number of variables and instruments
 - ✓ Ex. SEAC4RS Field Campaign: 3 aircraft, 53 research groups, 80+ instruments = 400 variables!
 - ✓ One variable may be measured by multiple instruments
- Heterogeneous data
 - ✓ Different types of data: scalar, vector, array, etc
 - ✓ Reported in different time scales and intervals
 - ✓ The same variable may be reported in different conditions
 - ✓ Data may not be reported for all flights

Complexities with Airborne Data (II)

- from a data system's perspective

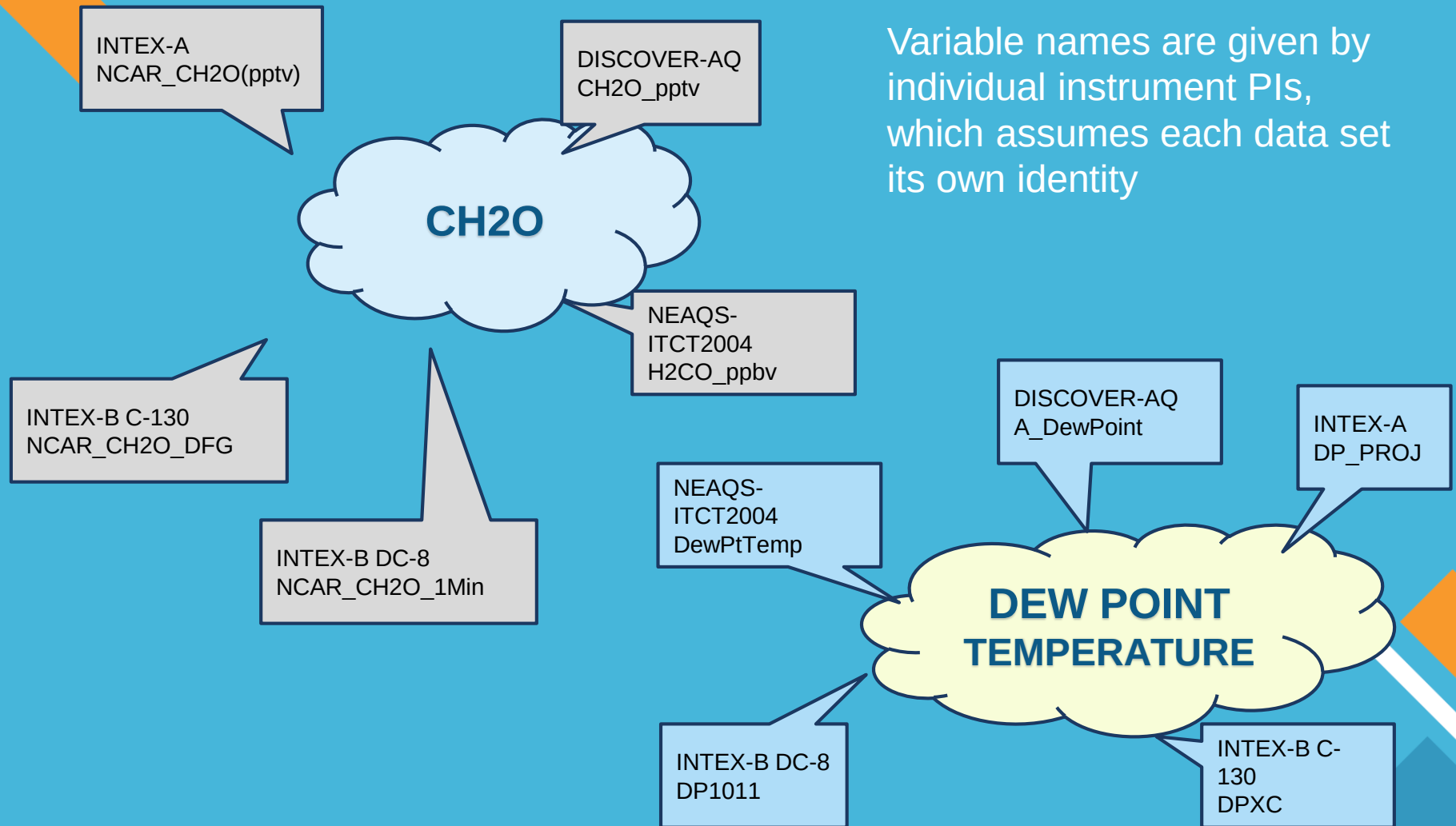
- Data format not rigorous enough so that the embedded metadata are machine comprehensible
- No consistent variable naming convention across studies
- Irregular spatial and temporal coverage
 - ✓ Flight path/sampling location cannot be readily expressed by straightforward mathematical representations, e.g., grids
 - ✓ The irregularity in flight days and path are the result of factors that cannot be controlled

NASA Data Management Challenges

- Lack of variable name standards
- Lack of detailed requirements for DAAC ingest process
 - Different DAAC have different ingest processes and somewhat different metadata requirements
 - Lack of detailed metadata requirements
- Current metadata standards somewhat inadequate for airborne observations
 - Some current metadata requirements are not applicable to airborne observations, while some critical data use information are not captured
- DAAC data discovery and distribution tools are inadequate for airborne observational data sets

Many issues can be resolved through discussion involving data users, e.g., modelers, data providers (i.e., instrument PIs), metadata experts, and data system developers.

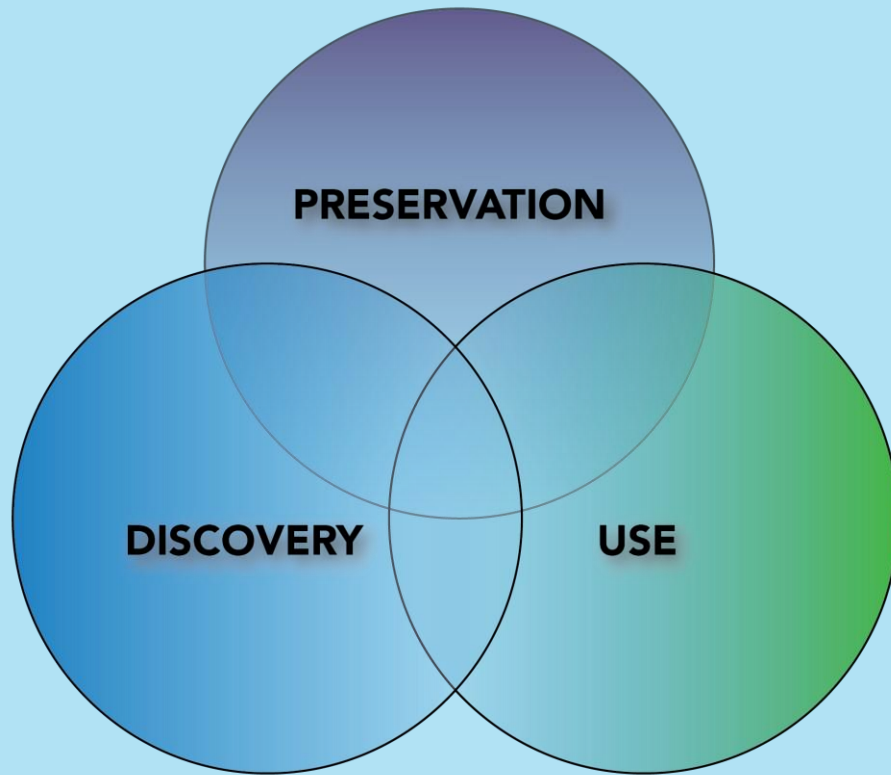
Variable Standard Name: Current Status



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:**
 - ✓ Discussions among data users, data providers, and data system developers to develop a community-acceptable list of variable 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

Metadata



- What is metadata?
 - Information about the data
- Metadata enables
 - ✓ Data system tools for data discovery, distribution, and subsetting
 - ✓ Record keeping
 - ✓ Proper data use

Metadata: Current NASA Standards

- NASA DAACs generally use Common Metadata Repository/Unified Metadata Model – Collection and Granule 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

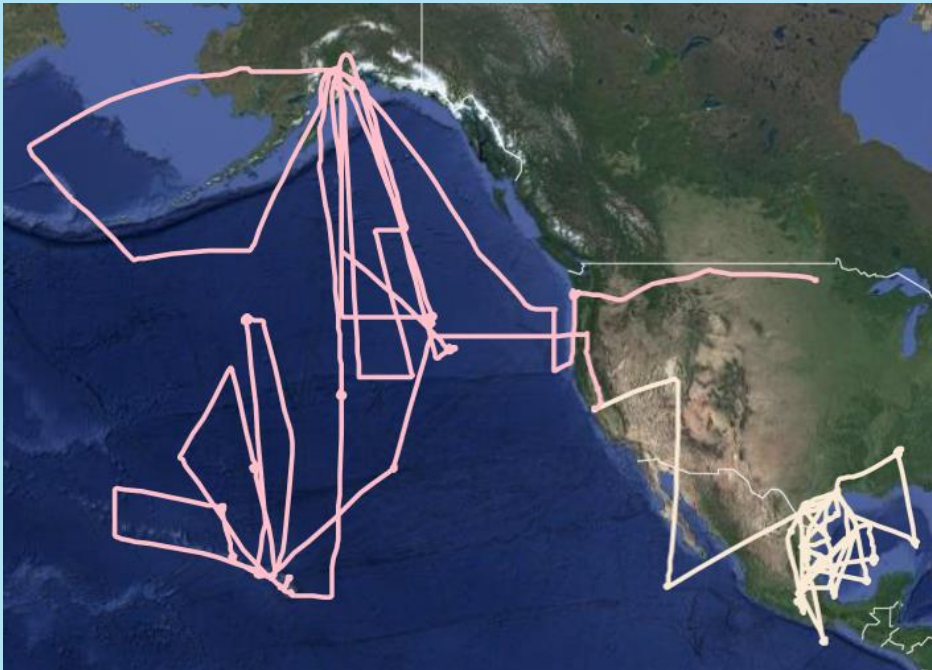
CMR/UMM-C	P	D	U
Metadata Info	√	√	
Data Identification	√	√	
Descriptive Keywords		√	
Distribution Info		√	
Temporal/Spatial Info	√	√	
Acquisition Info	√	√	

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

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

Metadata Challenges: Variable Description

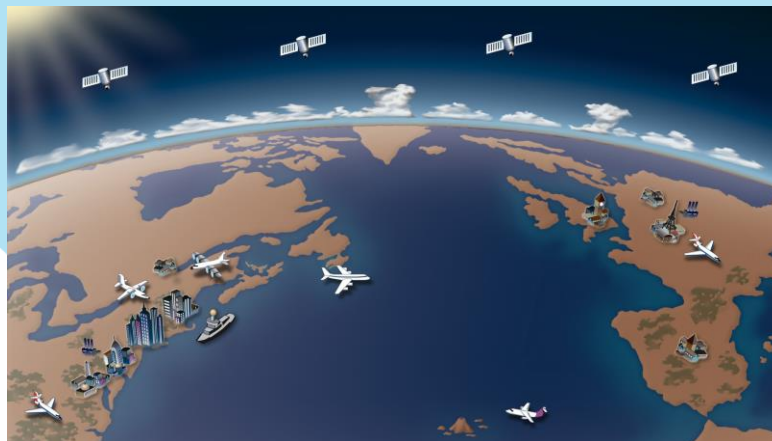
- For many in-situ observed variables, additional information are needed to properly use the data. For example:
 - A CO₂ measurement needs to be specified if the data are reported as dry mixing ratio or ambient mixing ratio
 - Aerosol mass concentrations need to be indicated if the data are reported 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 UMM-Variable:
 - 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

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
 - Not machine comprehensible
 - File Header metadata requirements are somewhat vague
- HDF and netCDF
 - Popular with modelers
 - Metadata expandable
 - Lack of standardized way to implement airborne metadata to address data discovery, data preservation, and data use issues

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 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 ESDIS Standard process working group



ICARTT Format V2.0

The ICARTT format was revised by the ESDIS ICARTT Format Refresh working group to address issues identified by data users, e.g., modelers, and data system developers. The ICARTT V2.0 was formally approved in 2017. The adopted changes were thoroughly discussed

- 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 including required keywords, order of the keywords, keyword/value pairs, and formatting for the keywords.

NASA Data DOIs

- Data DOIs (Digital Object Identification) are used to cite data sources
- NASA DAACs have been assigning DOIs to data sets for over 5 years, mostly for satellite data, to provide quick access to data sets used by journal publications
- There are also DOIs for airborne field studies:
 - INTEX-B DOI: 10.5067/Aircraft/INTEXB/Aerosol-TraceGas
 - DISCOVER-AQ DOI: 10.5067/Aircraft/DISCOVER-AQ/Aerosol-TraceGas
 - SEAC4RS DOI: 10.5067/Aircraft/SEAC4RS/Aerosol-TraceGas-Cloud
- The key issue is how to optimize between specificity and usability
 - DOIs assigned at the project level may be too broad
 - There would be over 100 DOIs if they assigned to each data product from each instrument PI. This would require expansion of the current journal review process to DOIs
 - There are attempts to assign DOIs to each manuscript. This may present a challenge to reflect the post-mission data updates
- A workable solution should be sought through discussion between researchers, journal editors, and data scientists

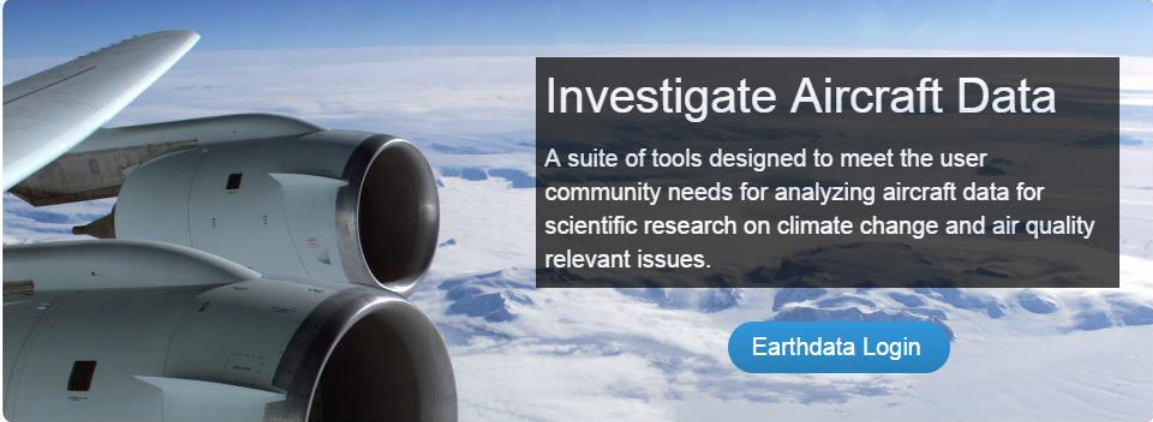
Toolsets for Airborne Data (TAD) Overview

tad.larc.nasa.gov

 Toolsets for Airborne Data

Documentation | Video Tutorial | FAQs | Mission Info | Relevant Links

Toolsets for Airborne Data (TAD) Web Application User Login Login



Investigate Aircraft Data

A suite of tools designed to meet the user community needs for analyzing aircraft data for scientific research on climate change and air quality relevant issues.

[Earthdata Login](#)

About The Project

This is a NASA ESDIS sponsored project aimed at providing user-friendly tools for accessing the data sets generated from airborne field studies for atmospheric composition measurements.

TAD allows users to discover data variables from multiple airborne studies and/or platforms. Based on their selections, users can output data products, including original data submitted by each measurement Principal Investigator (PI) and merged data, i.e., geolocated data from one or multiple measurement PIs.

Our Partners



Issues or Questions?

[View FAQs](#)

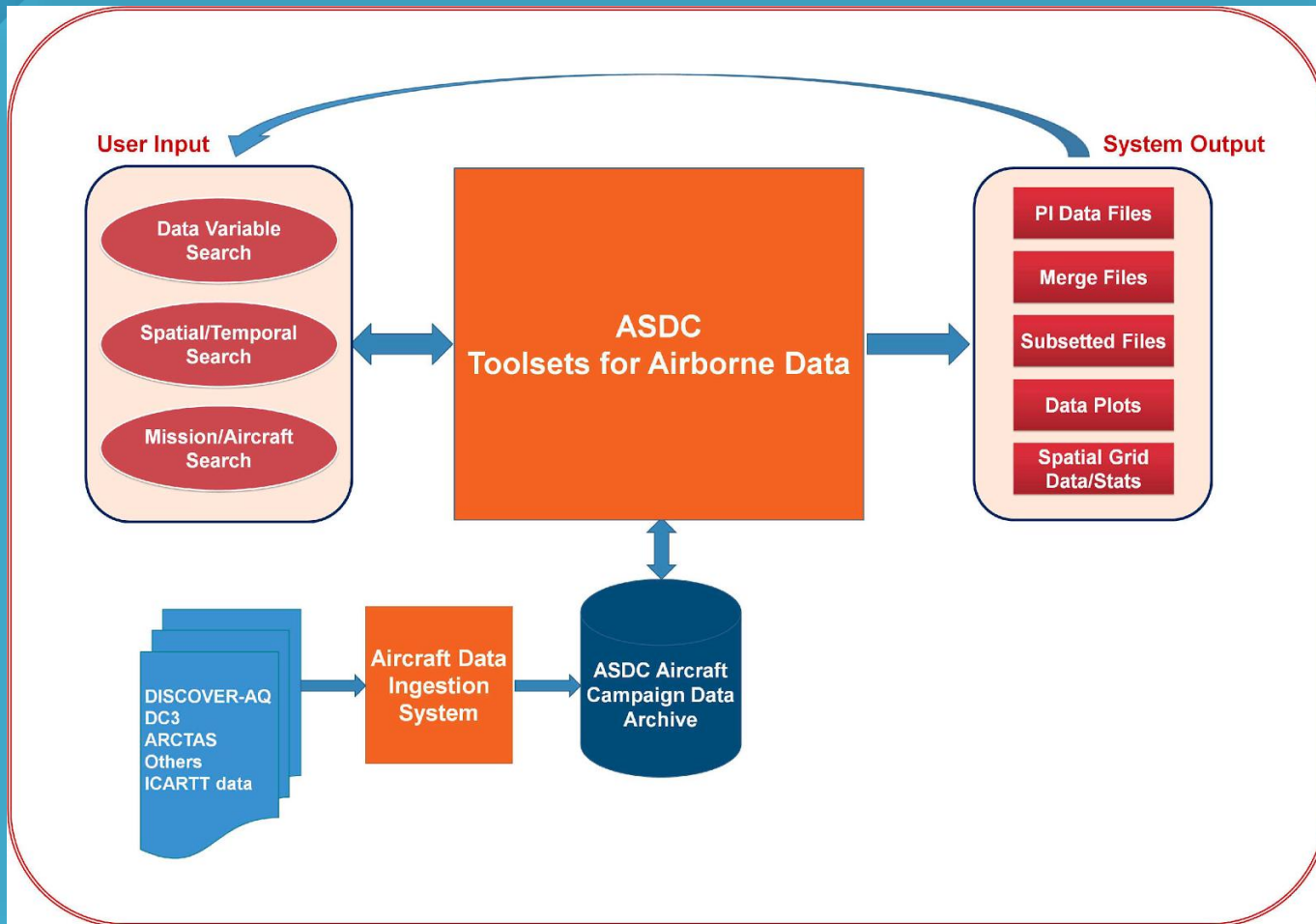


NASA Officials: Gao Chen, John M. Kusterer
Website Curators: Amanda Benson Early, Aubrey Beach
Last Updated: April 11, 2017

[NASA Privacy Statement, Disclaimer, and Accessibility Certification](#)
[Questions, Comments, or Feedback?](#)
[Site Credits](#)

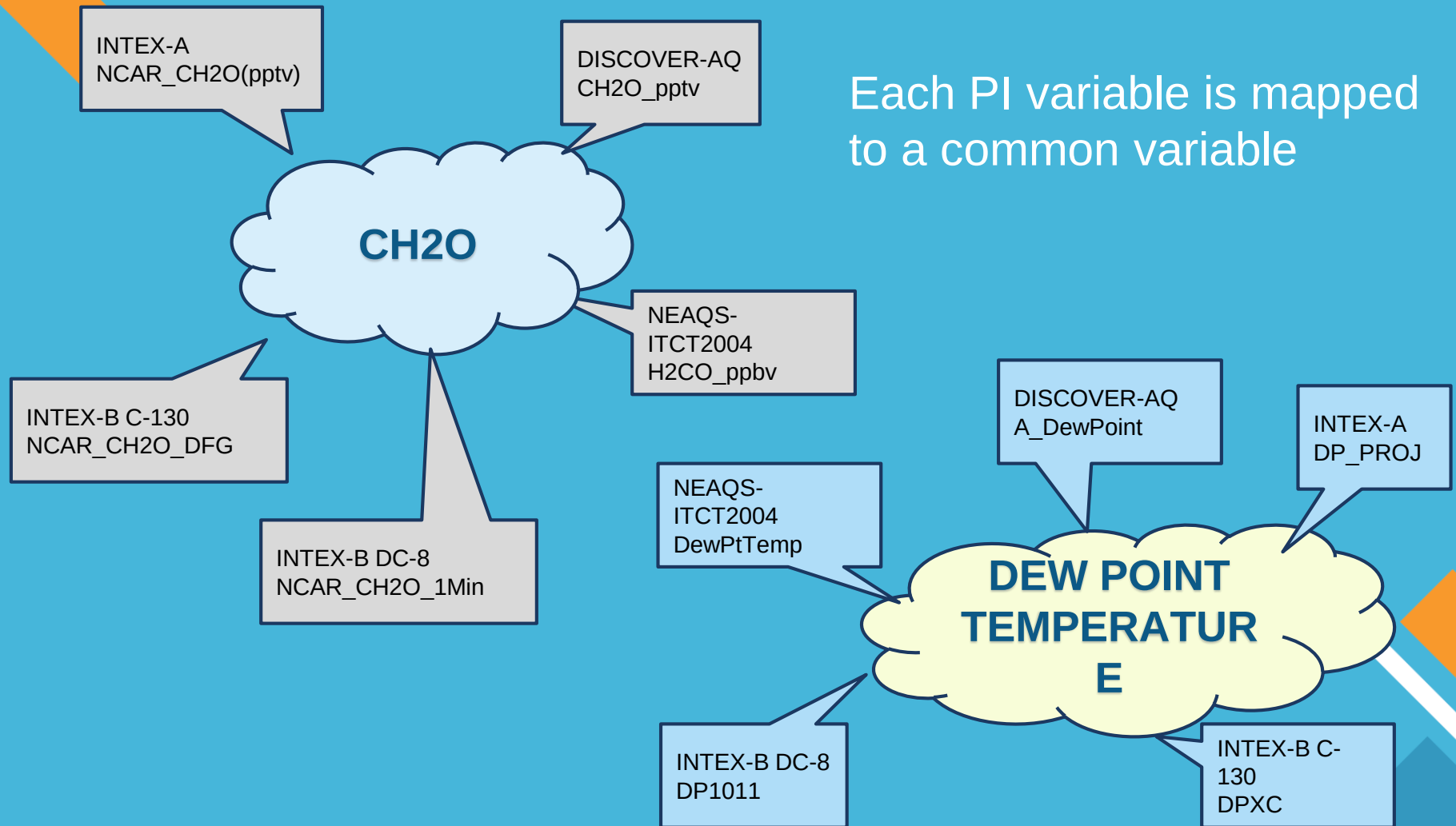
- TAD is designed to be a central web portal for airborne field study aiming to understand atmospheric compositions
- TAD currently handles only in-situ measurement data
- TAD allows users to download PI files and create merge files
- TAD is open to public, but registration (taking less than 3 min.) is needed for access

TAD – Motivation and Design



The spatial and temporal subsetting features are currently under development

Common Naming System to enable across study data search



User Interface Search Function

Select Variables

Create Merge Data

Request PI Data

Follow the steps below to select your desired data variable(s) based on scientific parameter, mission/platform, or date range. Once selected, choose Create Merge Data to receive a weighted average of requested data merged into a chosen time base or Request PI Data to receive raw data files based on the requested variables.

Step 1: Basic Search

Enter a Keyword...



Search by Parameter

Aerosol

Search by Mission

DISCOVERAQ MD

Search by Platform

P3B

Search by Date Range

07/01/2011



07/29/2011



Step 2: Refine Results

- ☒ Aerosol
 - ☒ Composition
 - ☒ Microphysical Properties
 - ☒ Optical Properties
 - ☐ Cloud Properties
 - ☐ J values
 - ☐ Meteorological/Navigational
 - ☐ Radiation Variable
 - ☐ Trace Gases

Step 3: Choose Variables

☒	Common Name	Mission	Platform	Instrument	Definition
☐	AE_Absrb	DISCOVERAQ MD	P3B	Derived	Aerosol Particle Angstrom Exponent For A...
☒	AE_Scattering	DISCOVERAQ MD	P3B	Derived	Aerosol Particle Angstrom Exponent For S...
☐	Aer_Absrb_blue	DISCOVERAQ MD	P3B	PSAP	Aerosol Particle Absorption Coefficient At ...
☐	Aer_Absrb_green	DISCOVERAQ MD	P3B	PSAP	Aerosol Particle Absorption Coefficient At ...
☒	Aer_Absrb_red	DISCOVERAQ MD	P3B	PSAP	Aerosol Particle Absorption Coefficient At ...
☐	Aer_Ammonium	DISCOVERAQ MD	P3B	PILs-IC	Aerosol Ammonium Ion Concentration
☐	Aer_BC_Mass	DISCOVERAQ MD	P3B	SP2	Aerosol Black Carbon Mass Concentraion
☐	Aer_Calcium	DISCOVERAQ MD	P3B	PILs-IC	Aerosol Calcium Ion Concentration
☐	Aer_Chloride	DISCOVERAQ MD	P3B	PILs-IC	Aerosol Chloride Ion Concentration
☐	Aer_Extinction_green	DISCOVERAQ MD	P3B	Derived	Aerosol Particle Extinction Coefficient At G...

1 2 3 4 5

1-10 of 42

Step 4: Confirm Selection

PI Variable Name	PI	Mission	Platform	Units	Data ID	Instrument	Description
X1_2_3_trimethyl_benzene	Atlas, Elliot	NEAQS-ITCT2004	NP3	pptv	NEAQS-ITCT20...	WAS	X1_2_3_Trimethyl_Benzene Mixing Ratio
X1_2_4_trimethyl_benzene	Atlas, Elliot	NEAQS-ITCT2004	NP3	pptv	NEAQS-ITCT20...	WAS	X1_2_4_Trimethyl_Benzene Mixing Ratio

User Interface – Merge/Download

Select Variables

Create Merge Data

Request PI Data

Merge Data Variable(s)

Follow the steps below to order a weighted average of data based on the selected variables that has been merged into a chosen time base. The variables are separated into accordions based on each mission/platform and must be ordered individually. To learn more, select the information icon at each step.

NEAQS-ITCT2004/NP3 – 2 variable(s)

Refine Variable Selection ?

Remove	PI Variable Name	PI	Units	Data ID	Instrument	Output Name
X	X1_2_3_trimethyl_benzene	Atlas, Elliot	pptv	NEAQS-ITCT2004-VOCsWAS	WAS	X1_2_3_trimethyl_benzene_Atlas
X	X1_2_4_trimethyl_benzene	Atlas, Elliot	pptv	NEAQS-ITCT2004-VOCsWAS	WAS	X1_2_4_trimethyl_benzene_Atlas

Select Merge Time Base ?

Custom

Enter Merge Time Base: 60

Note: Custom time base must be specified in integer seconds.

Select Flight Dates ?

Select Start Date: Select End Date:

Note: Unless a temporal subset is set, time domain for the date merge is the entire mission.

Select Processing/Output Options

☒ Average Only

☐ Average, Standard Deviation, Number of Points ?

☐ Average, Standard Deviation, Number of Points (with LOD processing) ?

Note: All output files will include the weighted average of each variable, merged into the chosen time base. Select additional processing options above.

Create Merge File(s)

Merge NEAQS-ITCT2004/NP3 File(s)

DISCOVERAQ MD/P3B – 3 variable(s)

TAD: Current and Pending Data Sets

Airborne Study	Deployment Year	Funding Agency	Ingest Status
DISCOVER-AQ MD	2011	NASA	Complete
INTEX-A	2004	NASA	Complete
INTEX-B	2006	NASA	Complete
ARCTAS	2008	NASA	Complete
NEAQS-ITCT 2004	2004	NOAA	Complete
DISCOVER-AQ CA	2012	NASA	Complete
TexAQS 2006	2006	NOAA	Complete
ARCPAC	2008	NOAA	Complete
CalNEX	2010	NOAA	Complete
DC3	2012	NASA & NSF	On-going
DISCOVER-AQ TX	2013	NASA	Complete
SEAC4RS	2013	NASA	Pending
DISCOVER-AQ CO	2014	NASA	Complete
NAAMES	2015 -	NASA	Pending
ATom	2015 -	NASA	Pending

Looking Forward

NASA has recently established the “Airborne Data Management Group”, which will be involved in

- Development of community-accepted variable standard names
- Development of a coordinated metadata collection process
- Enhancement of the current metadata standards to better fit airborne observations
- Development of prototype new data distribution tools tailored to airborne use needs