ASIA-AQ Field Data Repository dataID Registration and Data Upload

Michael Shook, Gao Chen, Ali Aknan, and Morgan Silverman

Introduction

- https://www-air.larc.nasa.gov/missions/asia-aq/
- A repository for all ASIA-AQ relevant observational and ancillary data products and relevant documentation/reports to facilitate science team data exchange and data processing
- File types include binary (HDF/netCDF), text (ICARTT), and image (jpg/png) files
- Password-protected document and file sharing

ASIA-AQ Field Data Repository

Points of Contact

- **Field Repository (www-air.larc.nasa.gov)**
    - Michael Shook, NASA Langley Research Center, michael.a.shook@nasa.gov, 757-864-5793
    - Gao Chen, NASA Langley Research Center, gao.chen@nasa.gov, 757-759-5642 (cell)
    - Ali Aknan, AMA/NASA Langley Research Center, ali.a.aknan@nasa.gov (website and file scanner)
    - Morgan Silverman, AMA/NASA Langley Research Center, morgan.l.silverman@nasa.gov (standard name issues)

Data Submission Steps

- ## DataID Registration (one-time process):
  - dataID is part of the filename and will be used to organize PI files on the data repository
  - Links data files to PI (see file naming convention slide)
  - PI or file creator will need to first register dataID(s) before files can be submitted

- ## Data Submittal and Scanning:
  - File submission is through a scanning tool (FScan) for checking filenames and content
  - All incoming files are scanned:
    - ICARTT Files: file header, including keywords and data flags, as well as time stamps
    - HDF and netCDF files: data variable dimensions and attributes
    - Other files: file names
  - Support zipped multi-file upload
  - Script-based batch upload and download available

- Username/password: contact POCs

DataID Registration Page

The PI's data directory will be created from LastName.FirstName. Please enter PI name correctly. Each dataID represents a [separate] group of files in PI's data directory.

See file naming convention slide

PI Last Name : [...]

PI First Name : [ ]

Platform (LocationID) : NASA DC-8 Aircraft (DC8) [∨]

dataID: (max 45 chars) *Prefix with "asiaaq-"* e.g., asiaaq-CO2	Data Description: (max 380 chars) *Describe your measurments; e.g., Carbon Dioxide Mixing Ratio*	Instrument(s): (max 190 chars) *List Instruments; e.g., LI-COR 6252*
Reset — asiaaq-	Enter one description per line AND press Enter	Enter one instrument per line

[+ Add Another dataID] [- Remove Last Entry]

Upload Your Instrument(s) Description Document: [Choose File] No file chosen (Select your file) ✖

(NOTE: Any new file/document upload WILL OVERWRITE your previously uploaded document)
(If you have more than 1 file (document) to upload, please email the files to: gao.chen@nasa.gov and/or ali.a.aknan@nasa.gov. Thank you!)

☐ OVERWRITE my previous record (i.e., ALL previously registered dataIDs for this mission will be removed).

Link to PI website, instrument, experiment description, etc
[http://] *Optional: to display on LaRC Archive webpage*

Text describing PI experiment or measurements (e.g., NASA LaRC DIAL - Troposphere O3, Aerosols, and Clouds Profiles):
[ ] *Optional: to display on LaRC Archive webpage*

[**Submit**]

Current Registered dataIDs on the Server for ASIAAQ 2024

PI Name: Last.First	LocationID	Registered dataIDs
Edit APEL.ERIC (PI Link)	DC8	asiaaq-TOGA-VOCs [+ Show Description]
Edit WOLFE.GLENN	DC8	asiaaq-ISAF-CH2O [+ Show Description]

Registered DataID Example

	dataID: (max 45 chars) Prefix with "asiaaq-" e.g., asiaaq-CO2	Data Description: (max 380 chars) Describe your measurments; e.g., Carbon Dioxide Mixing Ratio	Instrument(s): (max 190 chars) List Instruments; e.g., LI-COR 6252
Reset	asiaaq-HSRL2	HSRL measurements of clouds and aerosols, as well as DIAL measurements of Ozone Concentration	HSRL2
Reset	asiaaq-HSRL2-images	Full flight images of the lidar measurements of clouds, aerosols, and ozone, plus raster-specific images within a flight, designated as Ln (n can be 1,2,3,or 4)	HSRL2
Reset	asiaaq-HSRL2-mlh	Aerosol derived mixed layer heights and mean backscatter and extinction within several layers from HSRL2	HSRL2
Reset	asiaaq-HSRL2-NearSurface	High horizontal resolution measurements of HSRL2 atmospheric aerosol and ozone with lower vertical resolution. Several combinations of horizontal and	HSRL2

PI Last Name : HAIR

PI First Name : JOHN

Platform (LocationID) : LaRC GIII Aircraft   (LARC-G3)

+ Add Another dataID   - Remove Last Entry

Upload Your Instrument(s) Description Document: Choose File | No file chosen   (Select your file)

(NOTE: Any new file/document upload WILL OVERWRITE your previously uploaded document)
(If you have more than 1 file (document) to upload, please email the files to: gao.chen@nasa.gov and/or ali.a.aknan@nasa.gov. Thank you!)

☐ OVERWRITE my previous record (i.e., ALL previously registered dataIDs for this mission will be removed).

Link to PI website, instrument, experiment description, etc

https://science.larc.nasa.gov/hsrl/   *Optional: to display on LaRC Archive webpage*

Text describing PI experiment or measurements (e.g., NASA LaRC DIAL - Troposphere O3, Aerosols, and Clouds Profiles):

NASA LaRC High Spectral Resolution Lidar , Generation 2   *Optional: to display on LaRC Archive webpage*

Submit

Data Upload (FScan) Page Example

SARPEAST File Scanning and Submittal

ICARTT File Format Document

IMPORTANT: In order to archive files, **dataID(s) must be registered first** because the PI Data Directory on the Server will be created from the registration.

| Help |

What to do: ○ Scan Only   ● Scan and Archive (see dataIDs note above)

File Upload : Choose File No file chosen   (Select your file)

☐ This is Final (i.e., NOT Field) data -- File Revision MUST be set to R0, R1, etc.

| How to upload multiple files in one step | | How to archive *.zip ICARTT files |

This application scans three (3) file types -- defined as ICARTT: FFI 1001, FFI 2110, and FFI 2310

Select Data Time-Interval Type : Start, Stop, and Mid-point ⌄

   Start, Stop, and Mid-point

   Constant Equal (1 or less); {Or, irregular (Satellite/Trajectory/Ground)}

Display FScan results (report) : ● Detailed   ○ Brief

Submit   Pressing this button will Scan *[and Archive]* the selected file(s).

Please scan your files on your machine first using **FScanBrowser (Windows)**

*** Your Browser should allow FScan to open a new Window (to display FScan's results) ***

By: Ali Aknan -- Ver 3.7.8 (June 02, 2022)

*** Scanning Failed. See Error Messages Below ***

Please scroll down to the end of this page to view the results.

NOTE: error messages are displayed in RED; warnings and other info in ORANGE

Function Selected: Scan and Archive
FScan Results Report: Detailed
Time-interval Type Selected: Start, Stop, and Mid-point
Filename Submitted: DCOTSS-MMS-1HZ_ER2_20210817_RA.ict
File Size: 3066538 bytes
File Content: application/octet-stream

Receiving file was successful - now onto scanning ...

✔ **ALL DONE.**

Please scroll down to the end of this page to view the results.

NOTE: error messages are displayed in RED; warnings and other info in ORANGE.

Function Selected: Scan and Archive
FScan Results Report: Detailed
Time-interval Type Selected: Start, Stop, and Mid-point
Filename Submitted: ACTIVATE-LARGE-SMPS_HU25_20220111_RA_L2.ict
File Size: 55987 bytes
File Content: application/octet-stream

Receiving file was successful - now onto scanning ...

Note: results open on a separate page, may need to change browser permissions

ICARTT File Naming Convention for ASIA-AQ

DataID_LocationID_YYYYMMDD_R# [_Description].extension

- **DataID**: a short description of measured parameter/species, instrument, or model prefixed by "ASIAAQ-"

- **LocationID**: a controlled identifier of data platform, provided on the dataID registration website in a drop-down box.
  - ASIA-AQ locationIDs: DC8, LARC-G3, KINGAIR-1900D, KINGAIR-G90GT, MERGE, GROUND (for ASIA-AQ ground sites), AQ-MONITORING (for local air quality monitoring sites), MODEL, ANALYSIS, SATELLITE, TRAJECTORY, and OTHER

- **YYYYMMDD**: **UTC** date of takeoff for flight data or the beginning of the measurement for ground sites. Note: KST = UTC+9; PHST, MYT = UTC+8; ICT = UTC+7

- **R#**: Revision identifier. Typically, RA, RB, RC, … for field data and R0, R1, R2, … for the publication quality data. Note: archived files cannot be overwritten, *only replaced with subsequent revisions*

- **Description**: optional additional description of the file if necessary

- **Extension**: *.ict, *.nc, *.cdf, *.hdf, *.h4, *.h5, *.hdf4, *.hdf5, *.he5, *.kmz, *.kml, *.htm, *.html, *.txt, *.jpg, *.jpeg, *.gif, *.png, *.bmp, *.pdf, *.xls, *.xlsx, *.doc, *.docx, *.ppt, *.pptx

- **Extension upon request:** *.zip, *.tar, *.gz, or others

- The underscore, "_", is used ONLY to separate the different fields of the filename

- Examples: the filename for HSRL cloud and aerosol observations on January 31st, 2024:
  - ASIAAQ-HSRL_DC8_20240131_RA.h5 (for draft/field data)
  - ASIAAQ-HSRL_DC8_20240131_R0.h5 (for publication quality data)

ASIA-AQ Data Submission Schedule

Mission Phase	Data Type	Submission Deadline	Access Control
Field Deployment	Field Data	24 hour after each flight or cal. day	Science team and Partners
Post-Deployment	Publication-quality or "Final" Data	October 1, 2024	Public

- Field data submission deadline may vary depending on field operation constraints
- The final data should be of publication quality and time synced to the time standard for each platform

ASIA-AQ Data Format Requirements

- The ASIA-AQ data will conform to **ICARTT, netCDF, or HDF** format standards. This supports NASA's Open-Source Science and Open Data initiatives by making ASIA-AQ data FAIR

- All in-situ measurements are required to report data in ICARTT format (http://www-air.larc.nasa.gov/missions/etc/IcarttDataFormat.htm)

- ICARTT files will be scanned to ensure compliance with the format requirements

- HDF and netCDF files should be as CF-compliant as possible, i.e., having all required global and variable attributes and properly dimensioned data variables (template URL to be posted)

- In-field and remote assistance will be available to the science team to troubleshoot file format and submission issues

ASIA-AQ Data Reporting Best Practices

- Use the same number, names, and order of variables throughout the mission for files within the same dataID and revision. This prevents issues with the online merge tool, and FScan now checks the variable list between files

- Measurement Time Reporting:
  - Fixed variable name(s): Time_Start, Time_Stop, and Time_Mid
  - Report start, stop, and mid times if integration interval larger than 1 sec
  - Can use one time stamp (e.g., Time_Start or Time_Stop) for data at ≥ 1 Hz

- Use file header (e.g., DATA_INFO) or metadata to indicate whether the measurement time is synced to the time standard

- Trace gases: Indicate whether measurement is reported in dry or ambient condition

- Report absolute concentrations and aerosol extensive properties at STP: 273.15K and 1013.25 hPa (i.e., 0°C and 1atm)

- Use required attributes for HDF and netCDF files and properly dimension data variables

- Variable short name should not start with a number or contain "-"

- Recommend standard unit notation: WMO Codes Registry : wmdr/unit

ACVSNC Variable Standard Names

- Atmospheric Composition Variable Standard Name Convention (ACVSNC) is a NASA Earth Science Data Systems convention, intended to make data more findable and interoperable, and (re)usable:

  https://www-air.larc.nasa.gov/missions/etc/AtmosphericCompositionVariableStandardNames.pdf

- Constructed from controlled vocabulary

- Tags, **NOT** short names or variable names:

  - **For ICARTT files: short name, unit, *standard name,* long name**

    CH2O_pptv, pptv, *Gas_CH2O_insitu_S_AVMR,* mixing ratio by volume

    CH2O_LOD_pptv, pptv, *Gas_CH2O_insitu_S_AVMR,* Limit of Detection

    NOy_pptv, pptv, *Gas_NOyasNO_insitu_M_AVMR,* Total Reactive Nitrogen Mixing Ratio

    Sc700_total, Mm-1, *AER_Scattering_insitu_red_RHd_Total_AMB,* Dry Scattering at 700nm (Total Aerosols)

  - **For netCDF and HDF files: use ACVSN_standard_name attribute**

- Contact Morgan, Gao, or Michael for questions or adding new standard names

ASIA-AQ Science Data Policy

All participants are requested to accept the following responsibilities:

➢ Submit data in ICARTT, netCDF or HDF format no later than the deadlines

➢ If unexpected events lead to any delay in data submission, the PI is required to notify the project leadership as soon as issues are known

➢ **Publication-quality or "final" data should be submitted to the archive prior to any presentation at scientific conferences (e.g. AGU, AMS) or manuscript preparation, unless explicit authorization is obtained from the program managers**

➢ All aircraft measurements from a common platform should be synchronized to science team pre-agreed time standard

➢ Consult with PIs when using their data in conference/data workshop presentations and/or manuscript

➢ Invite PIs of any data used to be co-authors (particularly during post-deployment research phase)

➢ PIs should be available to answer questions about their data

ASIA-AQ Data Merge

- ASIA-AQ science team can use the "Custom Data Merging Tool" for DC-8 merges
- The project merge will be made and updated by the field repository staff when significant amount data are submitted or updated
- The "Custom Data Merging Tool" will allow science team member to make merges from ICARTT data files submitted to the repository with options to choose variables, merge time scale, and flight(s)
- Contact POCs for questions or requests