Data Management Plan for SEAC^{4}RS Airborne Field Study

Official version, May 28, 2013

Introduction

Studies of Emissions and Atmospheric Composition, Cloud, Climate Coupling by Regional Survey (SEAC^{4}RS) is a NASA airborne field study, which involves NASA DC-8, NASA ER-2, and SPEC Learjet aircraft as well as a suite of ground sites, including LIDAR and Ozonesonde measurements. To achieve the broad SEAC^{4}RS science objectives, a cohesive data management plan is necessary to promote collaborations across the various components of the field study and to maximize the scientific value of the airborne and ground-based observations. This document is intended to provide a framework through which the observational 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).

The SEAC^{4}RS data will undergo three stages within the project-lifecycle: field data, preliminary data, and final data. The field data is generated during the field deployment and is primarily used to measure progress in achieving the science goals. The preliminary data incorporates possible adjustments due to post-deployment instrument calibration/characterization, data synchronization, and the QA/QC process. The final step of the QA/QC process involves integrated data processing and analysis, which may reveal issues requiring reevaluation of preliminary data products. The final data is 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 main goal of the SEAC^{4}RS 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 project overall science objectives.

Data Repository

A data repository will be established for the SEAC^{4}RS field study at http://www-air.larc.nasa.gov before July 20, 2013. The Airborne Science Data for Atmospheric Composition (ASD-AC) group at NASA Langley Research Center will be responsible for maintaining the data repository. As the project progresses, the data repository 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 final data and associated documentation will be transferred to the data archive at the ASDC. The data submission schedule is given in the "Data Submission Deadline" section. The data repository will host data from the NASA DC-8, NASA ER-2, and SPEC Learjet aircraft as well as surface-based observations, including ozonesondes. The only exception to this is for the three satellite prototype instruments (AirMSPI, e-MAS, and RSP) onboard the ER-2, as they will work with their assigned DAAC (Distributed Active Archive Center) in order to simulate the data reduction process as if the instruments were making the measurement aboard a satellite. In addition to observational

data, the data repository will also hold all project-funded model results, satellite data products, meteorological forecasts, and back-trajectory calculations. Links will be provided for accessing the AERONET and MPL data archives.

The field data and preliminary data will be open only to SEAC^{4}RS science team and collaborators. The data download from the repository will be controlled by a common username and password. The field and preliminary data can be made available to those outside of the SEAC^{4}RS community per request to the project leadership. The final data will be open to public and will be transferred to LaRC ASDC. Data revisions during and after the project life cycle will be tracked through data file revision numbers as required by ICARTT file naming convention.

Data Archive Structure and DataID: The SEAC^{4}RS data repository will be constructed with a three-tier directory structure. The top, root level tier identifies the mission name: 'SEAC^{4}RS'. The second level tier is based on the platforms (for example, 'NASA DC8', or 'NASA ER2') 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 of file type. Otherwise, the system will not recognize their files as valid data inputs.*** The "dataID" is an identifier of the data source, which is typically prefixed by the project name and followed by 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, "DISCOVERAQ-DLH" was used as a "dataID" for diode laser hygrometer measurements of water vapor data during the DISCOVER-AQ study. For the SEAC^{4}RS campaigns, all "dataIDs" will be prefixed with "SEAC4RS-".

Data Submission Deadline

The data submission deadline, summarized in the table below, is designed to facilitate collaborative research for achieving the overall mission science objectives.

Mission Study Phases	Data Type	Submission Deadline	Access Control
Field Deployment	Field data	24 hour after each flight	Science team & Partners
Post-Deployment	Preliminary Data	April 15, 2014	Science team & Partners
Public	Final Data	October 15, 2014	Public

During the Field Phase, the 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 data will be deleted when the preliminary data are delivered (April 15, 2014). The preliminary data will be removed following the delivery of the final data (October 15, 2014). Data from prototype satellite instruments (i.e., AirMSPI, e-MAS, and RSP) will follow their internal data reduction cycles and will deliver the final data products to the assigned DAAC by October 15, 2014.

The preliminary data, due approximately 6 months after the field deployments, is primarily used for integrated processing and analysis, which serves as an important step toward finalizing the observational data. The final data will be made publicly available on October 15, 2014 through the data repository and also transferred to the archive at LaRC ASDC.

Data Format Requirements

SEAC^{4}RS 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 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 SEAC^{4}RS observational data must be reported with universal time (UT) for the time record. In addition, the SEAC^{4}RS project has a specific file naming convention to identify the airborne campaign, i.e., the file names will be prefixed with "SEAC4RS-" (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 troubleshoot issues in generating and/or submitting ICARTT files.

The SEAC^{4}RS 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, SEAC^{4}RS 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 appropriate directory. UT should also be used in HDF files for reporting time of the observations. It is also required that the PI should clearly indicate the

measurement/integration period by label the time stamps as start, stop and/or mid time. Finally, it is noted here that the ER-2 satellite prototype instruments, i.e., AirMSPI, e-MAS, and RSP should follow the formats from their agreements with the assigned DAACs.

Specific SEAC^{4}RS 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 to 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 ER-2	Primary: UAS Ozone (Unmanned Aerial System Ozone Instrument) Secondary: MMS (Meteorological Measurement System)
NASA DC-8	DLH (Diode Laser Hygrometer)
SPEC Learjet	GPS Time

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. It is recommended that the SEAC^{4}RS airborne study adopt a consistent naming convention for the short variable names and require long variable names along with the short names. This recommendation is intended 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 will be distributed to the 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 will be provided in another spreadsheet. Both spreadsheets will be also posted on the data archive website. The SEAC^{4}RS PIs should choose, if applicable, 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 a "-" (dash) 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 SEAC^{4}RS 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 SEAC^{4}RS 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 SEAC^{4}RS aerosol measurement PIs have reached a 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 SEAC^{4}RS, 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 SEAC^{4}RS participants are required to accept the following responsibilities:

- Submit data in ICARTT format no later than the specified 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.
- **Final data should be submitted to the archive prior to any presentation at scientific conferences (e.g. AGU, AMS, and AAAR) or manuscript preparation, unless explicit authorization is obtained from the project leadership.**
- All aircraft measurements from a common platform should be synchronized to science team pre-agreed time standard, e.g. DLH for DC-8
- Consult with PIs when using their data in conference/data workshop presentations and/or manuscripts.
- Consider inviting PIs of any data used to be co-authors (particularly during post-deployment research phase).
- PIs shall be available to answer questions about their data after submission.

During the Post-Deployment research phase, all SEAC^{4}RS investigators are encouraged to share data and collaborate with all groups associated with SEAC^{4}RS. Such data sharing must respect relevant guidelines. Component groups of SEAC4RS may also elect to share data or collaborate with groups outside the SEAC^{4}RS communities. Such data sharing with third parties will be arbitrated by the both project leadership and the relevant component group and will respect the protected status of data from the other component groups.

Acknowledgement Statements

When any of the SEAC^{4}RS data are used in a publication, an acknowledgement statement should be included, recognizing the efforts from the science team and the funding program and agency.

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 SEAC^{4}RS archive. The merge files will be made available at the data repositories. The merge files will be updated throughout the project lifecycle as the PI data files are revised.

Data Manager

The SEAC^{4}RS Data Manager will monitor the data submission status in accord with the data submission timeline. The data manager will also coordinate the efforts to support implementation of ICARTT format and the production of the data merge files.

Data Manager Contact Information:

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

Appendix A SEAC4RS data file naming convention:

dataID_locationID_YYYYMMDD_R#.extension

The only allowed characters are: A-Z 0-9_.- (that is, upper case alphanumeric, underscore, period, and hyphen). The use of the underscore character is restricted by the ICARTT format naming convention and may only be used to separate fields, as shown above. Fields are described as follows:

dataID: an identifier of measured parameter/species, instrument, or model (e.g., O3; NxOy; and PTRMS). For SEAC^{4}RS data files, the PIs are required to use "SEAC4RS-" as prefixes for their DataIDs, i.e., SEAC4R4-O3 and SEAC4RS-NxOy.

locationID: an identifier of airborne platform or ground station, e.g., 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 takeoff)

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., R0, R1, R2 … 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 September, 1, 2013 flight may be: SEAC4RS-DLH-H2O_DC8_20130901_RA.ict (for field data) or SEAC4RS-DLH-H2O_DC8_20130901_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 will need to be given as a minimum. Ideally, precision and accuracy will be provided explicitly. The confidence level associated with the reported uncertainties will also need to be 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.

Limit of Detection Information: definition of the upper and lower limits of the instrument or measurement technique (or N/A if not applicable), as well as flag codes for when measurements were outside of those bounds (separate flag codes should be provided for above the upper limit of detection and below the lower limit of detection).

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.