CHACHA Data Management Plan

Stony Brook University (SBU), University of Michigan (UM), Penn State University (PSU), Purdue University (PU), University of Alaska-Fairbanks (UAF) and University at Albany (SUNY) PIs will comply with the data policies of the NSF Division of Polar Programs and the Division of Atmospheric and Geospace Sciences in order to make the results of this research visible, accessible, and independently understandable to general users, free of charge or at minimal cost, and in a timely manner, with appropriate and helpful metadata. The CHACHA co-PIs will work together to consolidate and archive the project data to optimize its usability.

1. *Expected Data*

Most of the data for this project will be semi-continuous data acquired from the two aircraft. All clocks related to data acquisition will be GPS-time synced prior to each flight. Within 24 hours following each of the flights, raw datafiles and quicklook plots will be uploaded to a data repository managed at SBU and circulated among collaborators. After the flights are completed, L0 (uncalibrated) data will be converted to L1 (calibrated) data, and then error-checked. Final L1 data sets will be recorded in sharable files in a common format, e.g. UTC time; Latitude; Longitude; GPS altitude; concentration or physical parameter. As described in the proposal, this will include data for a wide variety of measurements, including temperature, pressure, radiation, O_3, H_2O, CO_2, CH_4, NO_2, Cl_2, Br_2, BrCl, BrO, ClO, $ClNO_2$, N_2O_5, HO_2NO_2, aerosol number concentration and size distribution, cloud particle size distributions, and a number of other lower frequency measurements. Some data products, especially the trace halogen gas and off-line aerosol composition measurements, are more complex, and will require substantial post-processing, e.g. CIMS, PILS-IC, CCSEM-EDX, and AMAX-DOAS data. All data files will have appropriate associated metadata files.

Flight science crew members will maintain electronic flight logs (to contain notes on each day's flight plan and any deviations, and visual observations about the weather, surface conditions, instrument performance, etc.), which will be stored and remain accessible, as accessory files to the data files. All flight data will be backed up daily to two independent storage devices, one of which will be cloud-based. All data processing steps will be documented by the student or scientist responsible for each individual instrument. King Air data for standard flight instruments will be available via netCDF files, for all processed data. The file formats follow the NCAR-RAF/nimbus conventions (https://archive.eol.ucar.edu/raf/ Software/netCDF.html). Software for viewing files is available from EOL (https://www.eol.ucar.edu/software-downloads). UWYO follows NSF/LAOF and NCAR/EOL guidelines/policies regarding accessibility to data. UWYO archives the King Air standard data on the UWKA website with a public link to the data, with ready public access. The UWYO group will produce Digital Object Identifiers (DOIs) for the King Air standard data sets to ensure the data is both publicly accessible and describable/discoverable to the broader community. It also provides data users a way to reference the dataset used in publications.

2. Archive Data Format

Within 6 months after each deployment (depending on the analysis complexity, e.g. for AMAX-DOAS, CIMS), revisions of individual instrument datasets will be uploaded to the data repository, after initial quality control measures (calibrations and internal diagnostics) have been applied. Offline IC and CCSEM-EDX data will be added, as available, with a goal of 1 year for most data. Within 1 year of each deployment, a master data file (with meta data), with measurements time-synced to a master clock, will be uploaded to the data repository (at the Stony Brook Academic Commons), to enable further quality control measures on each of the measurements. A data specialist at Stony Brook will assemble QC-ed working data sets for common analysis. L1 and L2 data products will be produced, e.g. production of data sets that are segregated by surface characterization. Archived data, in ASCII format, will be accompanied by a "README" file which explains the nature and objectives for each flight, the instrument and operator producing the data. Appropriate meta-data will be included with each file to

specify project, location, time, PI, operator, last modification date, and any details about the data pertinent to interpretation. The metadata for final archived data will include information about uncertainty estimates. Data from all flights and off-line sample analyses will be preliminarily stored at a project archive through the Stony Brook University Academic Commons system (https://commons.library.stonybrook.edu/; which has a longevity guarantee) with password-access for the co-PIs during the data processing period, and then for final archival, the project archive with finalized data will become open-access, where all project data will be stored, as well as to the arcticdata site, described below.

3. Access to Data

Final L1 instrument data, as well as the associated metadata, will be submitted to the NSF Arctic Data Center (http://arcticdata.io) within two years of completion of the field project in Utqiaġvik. The NSF Arctic Data Center provides data archival, preservation and access for projects funded by the Arctic section of NSF Polar Programs. All Arctic Data Center standards will be followed for data and metadata format and content. Shepson, Pratt, Fuentes, Lance, and Simpson will work with the Arctic Data Center data management tools to ensure that all metadata is accurate, understandable, and available. Data will be QC-ed and verified by the co-PIs prior to submission to these data archives.

4. Data Sharing Practices and Policies

Public access to the data will be achieved by using the NSF Arctic Data Center archive to ensure reaching the widest audience of Arctic/atmospheric science researchers. Following detailed data quality control, data archival will be completed according to the NSF Division of Polar Programs policy of archival within two years of data collection or by the end of the award (whichever comes first). UWYO standard products from the King Air are publicly available once a final QA/QC is completed, typically less than 6 mos. following the end of the project (or earlier). Data requests from researchers will be contingent on preserving our right to publish these data first and to be considered in resulting publications by others using these data. The data submitted to the NSF Arctic Data Center will include a formal citation and digital object identifier (DOI) to facilitate tracking of re-use and to give fair credit to the project. All publications resulting from the CHACHA project will refer to the data availability from the archive, and the procedures for requesting/retrieving the data.

5. Archiving of Data

Physical aerosol and PILS samples will be stored until analysis, and if sample remains, these will be stored at UM for at least two years after project completion. Processed data (e.g. AMAX-DOAS profiles) and interpretation of the data will be disseminated through reports, conference posters and presentations, and journal publications. Data processing, creation of backups, and data protection will be handled in a systematic fashion making use of existing facilities at the co-PIs laboratories. Data used for publications (peer-review journal and conference proceedings) will be stored on long-term archives at the SBU, the co-PIs laboratories, as well as through the NSF Arctic Data Center. Data in the SBU Academic Commons are freely accessible online, regularly backed up, easily discoverable via search engines, assigned a permanent URL so they can always be found, and preserved for use into the future. Data are guaranteed to be preserved, backed up and available for as long as the university exits. Academic Commons is hosted by Bepress on a Digital Commons platform. Digital Commons has a robust system of fast, failover servers, on and off-site backups, third-party archival services, and automated system monitoring to ensure 99.9% uptime. Furthermore, Bepress has two engineers on call twenty-four hours a day, every day. They back up customers' repositories every 4 hours and store the data off-site with Iron Mountain, a trusted data storage facility.