7-30-2020

Collaborative Research: Network Cluster: Dust in the Critical Zone from the Great Basin to the Rocky Mountains

Janice Brahney
Utah State University, janice.brahney@usu.edu

Data Management Plan

Types of Data Produced

Physical Samples: Samples will include (***Projects 1-6***) sediment, soil, bedrock, dust, snow, vegetation, and surface water. Each sample will be associated with metadata including a unique sample ID, location, sampling date, collector name, sample processing information, and mass collected.

Analytical Data: Meteorological measurements (***Projects 1&4***), PI-SWERL analyses (***Project 1***), integrated soil erosion variables (***Project 1***), geochemical characterizations (***Projects 2-6***), mineralogical investigations (***Projects 1, 5***), water chemistry (***Project 3***), grain size measurements (***Projects 4&5***), snow property observations (***Project 4***), dust in snow concentrations (***Project 4***), visibility measurements (***Projects 1&4***), temperature datalogger files (***Project 5***), snow albedo and dust radiative forcing maps (***Project 4***), geophysical measurements (***Project 5***), data on microbial communities (***Project 3***), and biological proxy data (***Project 6***) will be primary outputs from the analytical components of the CTZC.

Computational Data: Modeling efforts will yield a variety of computational data including gridded dust emission fluxes for observed dust events (***Projects 1&2***), time series of dust transport to specified receptor sites (***Projects 1&2***), dust particle trajectories (***Projects 1&2***), future dust emission scenarios (***Project 2***), modeled dust radiative forcing (***Project 4***), snow energy balance modeling outputs (***Project 4***), and predicted nutrient fluxes from the SWAT model (***Project 6***).

Data and Metadata Standards

Whenever possible, analytical data will be converted from proprietary formats to universally accessible ASCII and .csv files. Visibility images from time-lapse cameras, and multi-spectral images from the UAV (***Projects 4&5***), will be stored in jpeg format. STILT particle trajectories and simulation parameter information will be packaged and saved in a compressed .rds file (serialized singe R object) with the naming convention YYYMMDDDHH_LONG_LATI_ZAGL_traj.rds. STILT source footprints will be packaged and saved in a compressed .nc file conforming to the Climate and Forecast (CF) metadata convention with the naming convention YYYMMDDHH_LONG_LATI_ZAGL_foot.nc. This object contains information about the model domain, the grid resolution, and footprint values. STILT receptor time-series data will be stored in ASCII format. STILT gridded dust emission fluxes will be stored in both ASCII and jpeg formats.

Metadata will be recorded with each data file and will follow the Dublin Core element set. To ensure quality of geochemical data, isotope and trace element data will be referenced to international standards: $^{87}Sr/^{86}Sr$ to NIST SRM 987, common Pb isotopes to NIST SRM 981, $\delta^{2}H$ and $\delta^{18}O$ to Vienna-SMOW, trace element concentrations to NIST SRM 1643e, and Hg concentrations to NIST 1641d.

Files related to modeling will be particularly voluminous and require their own dedicated storage plan. Accordingly, CMAQ dust simulation files, including related WRF input files, will be stored on BYU's CAEDM servers or BYU Office of Research Computing servers. Account access during the project will be coordinated through Prof. Adams. Post-project, access control will be transferred to CAEDM management. Some result files may also be stored in sub-awardee Virginia Tech's digital archive (VTechData; https://data.lib.vt.edu). Due to the large size of CMAQ output files (can be terabyte in size), only input control and intermediate WRF files will be permanently stored. Case-specific folders will contain case descriptor (readme.txt), WRF input (namelist.input), WRF meteorology (met*.nc), WRF gridding (geo_em*.nc), WRF land use and output (wrfout* binary), and CMAQ runscripts (csh) files. The case descriptor files will contain descriptions of the WRF and CMAQ code versions and weather/land use/dust emission conditions run for each case. The namelist.input file will allow public users to rerun WRF conditions if desired. The other files can be used with provided CMAQ run scripts to reproduce the CMAQ results. This approach will reduce file storage by one to two orders of magnitude while allowing interested users to reproduce detailed CMAQ results for their own use.

Policies for Access and Sharing

The data generated in this project will be backed up immediately and automatically to secure cloud servers. In addition, redundant backups will be made on external hard drives. During analysis, datafiles will be password protected, with access granted only to students/co-investigators. Data generated in this project will be included as supplemental documents in publications. All published articles will be deposited into NSF's Public Access repository (https://par.nsf.gov/).

Prior to archiving, data and any remaining archived samples will be considered for sharing with other researchers as requested. All interested sample and or data users will be required to sign a sample and data use contract requiring them to acknowledge the investigators and NSF in any resulting publication or presentation.

Policies and Provisions for Reuse and Redistribution

All data sets resulting from this research will be organized in standard non-proprietary formats to ensure ease of sharing, reuse, and redistribution. Any and all of the data produced, subject to potential delays for publication, will be available to any person. PIs will assist those who may make inquiries to gain access to data and reports. The web page for the CZTC will clearly describe who to contact and how to access data and any restrictions on access. All relevant data will be openly accessible as online supplemental materials associated with published journal articles.

Plans for Archiving and Preservation of Access

Sample remnants and samples not analyzed will be stored in the laboratories of the PIs following standard protocols.

To ensure effective archiving and facilitate future access, data generated by this CZTC will be stored in two different ways. First, data generated by the PIs in their individual projects will be archived in repositories relevant to that discipline. For ***Projects 1&2***, computational data, along with necessary metadata, will be archived in the USGS ScienceBase data repository (https://www.sciencebase.gov/). Mineralogical and geochemical data collected for soil, dust, and rock samples in ***All Projects*** will be archived in the Integrated Earth Data Applications EarthChem database (http://www.earthchem.org/). The HydroShare repository (https://www.hydroshare.org/) will be used for archiving hydrologic chemistry and microbial data (***Project 3***). For ***Project 6***, all metagenomic sequence data will be deposited at NCBI's Sequence Read Archive-SRA (https://www.ncbi.nlm.nih.gov/sra), and ecological data will be archived in Dryad (http://datadryad.org/) and DataOne (http://www.dataone.org/), as appropriate. In general, data will be uploaded to archives at the time of publication or two years following the completion of the project.

Archiving the data resulting from the various CZTC projects in discrete, relevant repositories will ensure that these data are visible and accessible in the main locations used by the scientific community. At the same time, it will be useful to have all data from this CZTC available in a single location to aid future researchers seeking data from this collaborative effort. Therefore, a complete copy of all data generated by the CZTC will be deposited in the Data@Middlebury repository, administered by Middlebury College Library staff, where it will be freely available online. Data in the repository will have standard Dublin Core metadata and include a registered Digital Object Identifier (DOI), to facilitate search, retrieval and association with published work. Digital Collections at Middlebury strives to be OAIS (ISO 14721:2003) compliant, to maximize accessibility, interoperability, and long-term preservation. Data submitted to the Data@Middlebury repository will become the responsibility of Middlebury College Library and will be retained for the life of Middlebury College unless removed to a satisfactory discipline specific repository. Digital archiving and preservation activities are guided by the Middlebury Digital Preservation Policy and administered by Middlebury College Library.