

Data Management Plan

Our project will provide for the preservation, documentation, and sharing of data collections, curriculum materials and other related research and education products using the guidance provided in NSF 11-060 OCE Data and Sample Policy. We will conform to the overall NSF philosophy of data management and dissemination, embodied in the NSF Award and Administration Guide (AAG) Chapter VI.D.4. The project will generate new data and incorporate existing data streams from ongoing projects. The specific data sets managed by this project include physical and biological variables sampled by the High Frequency Radar (HFR) network, glider AUVs, RHIB surveys, and animal borne instruments.

Physical Data: The data management for the physical data will be based on the considerable infrastructure already in place at RU, UD, and UAF to support HFR and glider operations as part of Integrated Ocean Observing System (IOOS). Three of the PIs on this project are also part of IOOS regional observing components in the Mid-Atlantic (MARACOOS) and Alaska (AOOS). This involvement will ensure that the unique Antarctic datasets (HFR and glider) will be managed in a way consistent with IOOS. This approach is based on a centralized data service that provides easy access to data and metadata.

HFR: The HFR data will be archived in a manner consistent with the specifications identified by the U.S. IOOS National HFR network within which RU, UD and UAF all participate. These methods have been developed as part of an operational HFR network supporting real-time search and rescue operations for the U.S. Coast Guard around the United States. At each remote site, the data will encompass raw level binary data that will be stored locally and transferred to a centralized server at RU when the sites are serviced at the beginning and end of each season. Throughout the deployment, real-time radial component files will be delivered to a centralized machine at Palmer Station and combined to generate total vector files. As with the spectra data, these files will be transferred and stored on a centralized server at Rutgers.

Gliders: For each deployment the complete dataset will be stored locally on the glider. In addition, a subset of the data files recorded by the glider in real-time will be transferred back to shore via the iridium satellite communication system. These files are then archived to a fileserver at Rutgers, where they are backed up daily. The raw data stream will be processed to be consistent with both the IOOS national glider plan and the NSF Ocean Observing Initiative (OOI) data management plans. Scientific (i.e., temperature, conductivity, depth, optics, and acoustics) parameters are merged with the glider navigational parameters (i.e., location, time) and are stored in organized data structures, which are saved to the RU fileserver in near real-time. Real-time glider health and deployment status will also be available on the internet at: <http://marine.rutgers.edu/cool/auvs>. This webpage will include plots of relevant scientific parameters (temperature, salinity, density, chlorophyll concentration, etc.) and maps showing the gliders path and intended waypoints. These processed datasets will be made available in near real-time in the trajectory NetCDF file format via the Thematic Real-time Environmental Data Distribution System (THREDDS). While the glider is on a mission, the real-time distributed data will be considered provisional until the complete dataset is quality controlled after recovery. Once the glider has been recovered, files containing the full datasets will be downloaded and the previous steps repeated, providing the end user with the complete scientific and navigational data streams.

Mooring and RHIB Surveys: Mooring data will be downloaded after recovery and immediately backed up on multiple hard drives. All physical data from the ACROBAT surveys will be downloaded after each survey and immediately backed up on multiple hard drives. All raw level binary data from the mooring and survey will be stored locally and transferred to centralized servers at both RU and UAF at the end of each season. Final processed and quality controlled data will be made available together with metadata at the end of the project at the national archive facility; National Centers for Environmental Information (NCEI) and via the

project website and data portal.

Model: Model output fields from the ROMS simulations (4d model physical variables such as velocity, temperature and salinity and model drifter tracks) will be in NetCDF format and will be compliant with the NetCDF Climate and Forecast (CF) 1.0 metadata conventions. All circulation model output will be stored at the Biological and Chemical Oceanography Data Management Office (BCO-DMO, <http://bco-dmo.org/data/>) at Woods Hole Oceanographic Institution. This service was created to help serve investigators funded by the NSF Geosciences Directorate (GEO) Division of Ocean Sciences (OCE) Biological and Chemical Oceanography Sections and Office of Polar Programs (OPP) Antarctic Sciences (ANT) Organisms & Ecosystems Program to archive and serve data generated through NSF-supported science. BCO-DMO will also archive all the data they manage at the appropriate national archive facility, such as NCEI and NGDC. Although BCO-DMO started as a data site for biological and chemical observations, but they have expanded their storage activities to include numerical model results. Model (final, quality-checked versions used in publications) data uploaded to BCO-DMO will have no restrictions on access other than those imposed by BCO-DMO. We will also archive for five years at ODU all numerical simulation data (in NetCDF) produced by this project and used in publications. After the first five years only the software configurations used to generate the numerical integrations (to insure their reproducibility) and figures will be maintained for at least another five years. ROMS code is open-source and freely available from the model developer.

Biological Data: All acoustic and net-tow data will be archived and managed by the PAL-LTER project and made available to the Convention for the Conservation of Antarctic Marine Living Resources (CCAMLR). CCAMLR uses acoustic data collected around Antarctica for the improved management of the Antarctic krill fishery. The PAL-LTER has a long history of data management and dissemination associated with data originating in the Western Antarctic Peninsula, including the Anvers Island region (see <http://pal.lternet.edu>). Archiving our acoustic data with the PAL-LTER makes strategic sense as it not only strengthens our connection to the project, but also ensures that our data are made available to the public and scientific communities for enhancing other lay and scientific endeavors (See attached letter of collaboration). Additionally, all chemical and biological oceanographic data collected by the science party will be submitted to the NSF-supported BCO-DMO.

Animal Borne: All data originating from the penguin tagging, including the referenced ancillary data, will be archived and served to the public and scientific communities through the PAL-LTER (<http://pal.lternet.edu>). This program already has well-established protocols for managing these types of data, and because of the program's long existence is now recognized nationally and internationally as an important archive of Western Antarctic Peninsula ecological data.

The PIs on this project have extensive experience managing and disseminating quality controlled datasets, specifically in relation to the technologies proposed as part of this field program. In addition, the overlap with large data aggregation programs like IOOS, OOI, and the PAL-LTER ensures that the same standards of data access and sharing will be applied to our project. Our data management will be carried out in a way to ensure that these datasets can be incorporated into the existing PAL-LTER methodology (<http://pal.lternet.edu>).