GRACE SABA: DATE MANAGEMENT AND SHARING PLAN

This project will provide for the preservation, documentation, and sharing of data collections, curriculum materials and other related research and education products in compliance with NSF policy, and I will conform to the overall NSF philosophy of data management and dissemination, embodied in the NSF Grant Proposal Guide (GPG) Chapter II.C.2.j. The project will generate new data and incorporate existing data streams from ongoing projects. The specific data sets that will be produced during this project include physical, chemical, and biological variables sampled during a glider AUV deployment and ship-based sampling for validation of the glider acoustic package. All significant findings from the proposed work will be promptly prepared and submitted for publication with authorship that accurately reflects the contributions of those involved.

Planning prior to fieldwork: Fieldwork is proposed for Year 1 in the Terra Nova Bay and adjacent shelf (western Ross Sea), Antarctica. Additionally, glider/sensor lab testing as well as glider deployments and ship-based fieldwork during glider deployments/recoveries in the Ross Sea are proposed for Year 1. Planning prior to the field season will be done with USAP logistics via teleconference or an in-person planning meeting in Year 1.

Types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project: 1) Description of the platform, the biological source and treatment protocols, description of overall experimental design, aim, and conclusions; 2) Sampling event log and data inventory; 3) Environmental (temperature, salinity from CTD casts during cross-calibration at glider deployment and recovery), biological (from ship-based tows and trawls from acoustic validation sampling: zooplankton biomass, size, and community composition; Antarctic silverfish biomass and size); 4) Data generated from transects with a Slocum Webb glider in the Ross Sea (salinity, temperature, fluorescence, backscatter, dissolved oxygen, acoustic data).

Plans for archiving data, samples, and other research products, and for preservation of access to them: The data management for this project will be based on the considerable infrastructure already in place at Rutgers to support glider operations as part of the Mid-Atlantic Regional Association Coastal Ocean Observation System (MARACOOS, http://www.maracoos.org) a regional component of the Integrated Ocean Observing System (IOOS). Dr. Saba serves as a co-PI on MARACOOS. This involvement will ensure that the unique datasets will be managed in a way consistent with IOOS. This approach is based on a centralized data service that provides easy open access to data and metadata.

  The project is a multidisciplinary effort, so several types of data and different data repositories will be involved in the data archiving effort as summarized below. These data repositories are well-established in their protocols for quality assurance, accessing and availability of data, and the specific details of data management are available through those centers.

  <u>Vessel-based data</u>: As directed by the U.S. Antarctic Program Data Center, discrete measurements collected during the collaborative research cruise in the Ross Sea (zooplankton biomass/size/community composition, silverfish biomass/size) will be submitted to OCB-DMO (http://bco-dmo.org/data/) which handles experimental data as well as *in situ* observations. BCO-DMO is already funded to provide data management for NSF PLR ANT projects at no extra cost to this proposal, and will archive all the data they manage at the appropriate national archive facility, such as the NODC and/or the National Geophysical Data Center (NGDC). I will submit all these data collected, after quality assurance as soon as possible but no later than two (2) years after the data are collected.

  <u>Glider Data</u>: Glider profiles conducted in the study area will consist of measurements of conductivity, temperature, depth, spectral backscatter, chlorophyll fluorescence, dissolved oxygen, and acoustic data. 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. Once the binary encoded files arrive on shore, they are converted to ascii text

using a set of unix utilities. 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 Rutgers fileserver in near real-time. Real-time glider health and deployment status will also be available on the internet at RUCOOL (http://marine.rutgers.edu/cool/auvs). This webpage will include plots of relevant scientific parameters and maps showing the gliders path and intended waypoints. These processed datasets will be made available in near real-time in the NetCDF file format via the Thematic Real-time Environmental Data Distribution System (THREDDS). Once on the THREDDS data server, the glider data is made available to a data asset map being run through the IOOS/MARACOOS program (http://assets.maracoos.org/). 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.

PI Saba has experience managing and disseminating quality controlled datasets, specifically with 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 ensure that the same standards of data access and sharing will be applied to our project. The Saba lab will design and host an initially password-protected share drive that will serve as a repository for the data generated from the vessel-based studies and glider transects (those data streams that are not delivered in real-time) which will be made available to the PI and associated team members for analysis. Public release of these data (to OCB-DMO, NODC, and other appropriate national data servers) will occur no later than 2 years after data are collected (as described above) and will likely coincide with the submission of the analyzed datasets to peer-reviewed journals.