

Data Management Plan

Summary of Approach: Data management activities include the collection and archival of project data sets, remote sensing data, and sample collection forms, as well as final synthesis and model outputs. We propose a coordinated and integrated approach that ensures the highest quality possible for all facets of the program and ensures the integrity of data and samples throughout the duration of the study. The objectives of data management are three-fold: 1) to ensure data integrity throughout the study, 2) to provide program management with the tools needed to monitor and guide study progress, and 3) to facilitate the dissemination of data to project scientists.

Types of data, samples, physical collections, software and other materials expected:

The research team expects to derive data and metadata from deployments of the surface and profiling autonomous vehicles (AVs) and from the two planned research cruises. The AVs will generate both near real-time and delayed-mode data streams. These streams will contain time series of: systems engineering data, 2D- or 3-D trajectories and measurements of temperature, conductivity, fluorescence, turbidity, turbulence and dissolved oxygen. Metadata includes information on parameter names and units, sensor descriptions, quality flags, geographic bounding box, contributor contact information and other project and platform level information.

Near Real-Time Data: Near real-time data are telemetered to shore at approximately 6-hr intervals. Due to power and bandwidth limitations, near real-time data are usually low-resolution subsets (~5%) of the full-body of data. Near real-time data have value in time critical uses and guard against a total data loss in the unlikely event that a vehicle is not recovered. These data will be made available to the project within minutes after receipt by the shore-side systems. Within hours, subsets of the data will be pushed to the IOOS National Glider Data Assembly Center (NGDAC) and passed to Global Telecommunications System for use by the world weather and modeling communities.

Delayed-Mode Data: Delayed-mode data are the full-resolution datasets downloaded from the AVs when they are recovered for servicing or at the end of the project. The full-body of scientific data will be processed into convenient forms for use by project participants and archiving.

Cruise Data: The proposed research plan calls for two cruises. Depending on the onboard environmental instrumentation available on the research vessel (expected to be the R/V Pelican), the expected data include: vertical profiles of temperature, conductivity, and dissolved oxygen; optically-measured parameters: fluorescence, transmissometry, optical backscatter, and photosynthetically active radiation; underway upper ocean currents (shipboard ADCP), temperature, salinity, and fluorescence; marine meteorological measurements and navigation logs. Data derived from chemical analysis of water samples

collected with Niskin bottles (nutrients, Winkler titration, pCO₂, etc.) will be assembled into a database suitable for archive. All data collected aboard ship will be submitted to the NOAA Center for Environmental Information for long-term archival within 12 months of collection.

Data and Metadata Standards: The expression of data and metadata are increasingly standardized. This increases usability and enables automated machine-to-machine exchanges. Standards are lists of what pieces of auxiliary information are needed to fully use the data, definitions of allowable words and terms and how values should be encoded. These standards are maintained by various communities of expertise and adopted by communities of practice. Standards applicable to oceanographic data are maintained by several entities known to the PIs (e.g., GRIIDC, NCEI and the IOOS NGDAC). The DM will be responsible for submitting and archiving data and metadata in conformance with these communities.

Units: SI units will be used to report parameter values except where the oceanographic community practice is otherwise.

Metadata: GRIIDC requires each dataset be accompanied by project-level metadata in an external file which follows the ISO 191** family of standards. This file will be created using the online editor provided by GRIIDC for this purpose. Cruise and delayed-mode AV data will follow the "NCEI NetCDF Templates v2.0" using the appropriate templates for AVs, profiles, and underway data. Near real-time data will follow the IOOS NGDAC V2.0 template.

Policies for accessing and sharing data: the team's default sharing policy is free open public access to all data and metadata collected by the project with a minimum of delay. The project Data Manager will provide access to project data through email, ftp, web services and web accessible folders (WAF). Public access may occur through websites maintained by the project, GRIIDC, IOOS NGDAC or NCEI.

Plans and Timelines for Archiving: The research team will archive the full-body of data and metadata with appropriate government data repositories and mark it as publically-available within 12 months of data collection for field and laboratory analysis, or 12 months following recovery for internally-recorded data. Near real-time data will be available immediately through GCOOS websites and in the case of profiling vehicle data, through the GTS and IOOS NGDAC servers. Cruise data will be submitted separately on a per-cruise basis. CTD data and other shipboard data (SCS, ADCP, met) will be archived according to the NSF R2R system. The full-body of data will be submitted to NCEI for long-term archiving through archiving agreements made between NCEI and the project team. Additionally, NCEI will automatically archive near real-time data submitted to the IOOS NGDAC at the close of a mission.