Data Management Plan

We recognize the benefits of freely sharing the data produced during this project with the scientific community and general public and of archiving data in a format that provides maximum interoperability between our collaboration efforts, supports modeling studies initiated by interdisciplinary groups, as well as provides support to the broader user community to ease their interpretation and analysis of the simulated data.

Data inventory

The observational part of this study will primarily use various datasets described in the above proposal. In general, most of these data sets are already archived in the public domain (i.e., BCO-DMO, NODC, AVISO etc.). Datasets may include, but are not limited to, time-varying regular or irregular gridded data (e.g., observational and re-analysis data). We will provide any interested parties instructions on how we accessed the data and all pertinent information about any post-processing of the data that was done.

The numerical modeling results will be available to any interested parties upon request. We will archive the model input files and output data files for all numerical runs and, with the exception of the initial testing runs, they will be made available.

General Data and metadata standards

Simulated output from models will use the following standards for their primary datasets:

NetCDF Files. NetCDF is a very widely used file format that has been specifically designed for storing scientific data, including model output. It is a flexible, self-describing format that bundles data together with any relevant metadata. It can be used to store virtually any kind of data including spatial grids, time-indexed grid stacks, time series, time-indexed profiles (e.g. ocean temperature vs. depth) and time-indexed "cube series". Due to the flexibility of the format, several standards have emerged that specify protocols for handling issues such as unstructured grids, names of physical quantities, inclusion of units, and so on. Among these is the so-called Climate and Forecast (CF) convention conformant, see: http://cfconventions.org/. The netCDF format has become so widely used within the scientific community that it is supported by virtually all analysis and visualization packages, including MatLab, IDL, Python, NCAR Command Language (NCL). In addition, APIs have been developed for virtually every programming language to simplify reading and writing data to netCDF files. There are many other supporting tools for netCDF such as NetCDF Markup Language (NcML) and an online tool for checking whether a netCDF file complies with the CF conventions.

CSV files and Multi-column Text Files: Some simple tabular data will be provided using ASCII column text files. These are human-readable and in an appropriate format that will be implemented for small size post-processed (synthesized) data.

Data Use, Privacy and Sharing Standards

The co-PIs will set up a research web page that will describe the main goals and approaches of the project. We will frequently update the research progress in a plain language for the general public along with several key figures. This research web site will be updated periodically by the

co-PIs and maintained free of charge by the WHOI Computer and Information Services (CIS) indefinitely including the backup service.

All simulated data produced and analyzed by this project will be archived to external disk drives. Ideally, we would like to keep these data available for the community for at least five years from the end of the grant, and longer if technically possible. We do not plan to place the data in a public repository, but will make any data available upon request. We will make every attempt to provide copies of the original or processed data in standard, nonproprietary formats (e.g., NetCDF files, ASCII files, TIFF image, etc).