Data management plan

1. **Types of data to be produced in the project**

   The main data produced by this project is listed in the following:

   (1) **Output from model simulations**, which will include both physical (T, S, velocities) and biogeochemical (tracer, rates, fluxes) variables. The raw output will be post-processed and analyzed for presentation at meetings and for publication in scientific journals. For model output from the main set of simulations (Table 1 in the proposal), standard model fields (e.g. physical and biogeochemical tracers, current velocities, and the main biogeochemical rates) at monthly resolution will be provided to, and made available through, BCO-DMO and NOAA NCEI in standard NetCDF format. In addition, the archival output fields will be stored at UCLA with daily-average values, suitable for use by others for independent analyses and for embedding local small-scale nests. These raw model outputs, which are generally too large and resource intensive to serve to an interactive portal, will be made available to the public upon request.

   (2) **New model code**, both in the form of the 1D model framework (including the optimization code), the ROMS physical model, and the 3D NitROMS implementation of the N cycle model as part of BEC-ROMS. The 1D advection-diffusion model code, which includes the revised N cycle model, will be documented and made available via a public GitHub page as soon as results are published, as is our practice, and no later than two years after the end of the project. This model is written in MATLAB language, includes in-line documentation, and can be run on a personal computer "out of the box". The 3D model code (NitROMS, which includes ROMS+BEC, plus the expanded N cycle) is presently maintained on a GitHub repository as part of Bianchi's group. It will also be released as soon as results are published, and no later than two years after the end of the project. Peer-reviewed articles based on the model will link to a public GitHub page, which will host the version of the code used to produce the results.

   (3) **Observational syntheses** from existing databases (for example gridded N cycle tracers) for model initialization and validation, as well as a compilation of N transformation rates and N cycle parameters from the literature. We will not attempt to serve existing data that are publicly available through data portals, but our in-house compilations will be documented and made available through BCO-DMO at time of publications of the results.

   (4) **A regional database** of biogeochemical measurements from the S. Monica Bay, collected during the undergraduate student field-trip cruises that form part of the Broader Impacts of the project. This database will be processed, documented and released via BCO-DMO on an annual basis, and no later than two years after the end of the project.

2. **Data standards and formats**

   Model output will be archived as self-contained NetCDF files, whose meta-data is readable via the NetCDF libraries with a wide range of software analysis tools. Data syntheses will be released either as NetCDF files, or as self-contained structures in MATLAB format. Model code will be released as FORTRAN (ROMS, NitROMS) and MATLAB code (1D model framework, processing optimization and analysis tools), and will include inline documentation and "how-to" tutorial files.

3. **Policies for data access and sharing**
   There are no confidential or private data expected in this study. The data syntheses, model code and processed model output will be stored on publicly accessible repositories (GitHub, BCO-DMO and NOAA NCEI) and linked by a project web page hosted at UCLA as soon as is feasible following processing or publication, and no later than two years after the end of the project. Due to the large size, raw model output will be archived on the PI's storage server and made available upon request.

4. **Policies for data access and sharing**
   All data, standard model output at monthly resolution, and raw model output will be released following the procedures indicated above, at time of publication of a peer-review article based on the data or model output, and no later than two years after the end of the project, whichever is soonest.

5. **Data archive plan**
   All data and model output will be archived on public repositories as outlined above, and backed-up on PI Bianchi's RAID server in the Department of Atmospheric and Oceanic Sciences at UCLA, where it will be preserved for as long as possible, but at least for two years following the end of the project period.