Data Management Plan

The proposed research builds on the foundation of (but is separate from) quarterly CalCOFI and CCE LTER sampling. Both programs have open data mandates, and thus have structured their information management and serving capabilities around open-data principles. All CalCOFI data are freely available (calcofi.org/data-policy.html). The data collected by the SIO CalCOFI technical group can be made available to users in preliminary form within ~ 1 to 10 days of collection by sending data from the ships ashore. The data are analyzed and quality controlled within approximately 6 months of each cruise and posted on the CalCOFI web site (www.calcofi.org) and in DataZoo (www.oceaninformatics.ucsd.edu/datazoo/), both publicly accessible data portals. The CalCOFI icthyoplankton data, curated by Southwest Fisheries Science Center (SWFSC), are available on a NOAA ERDDAP server (http://www.calcofi.org/data/erddap.html). CCE LTER also curates and serves data on the DataZoo platform, and all datasets (not including associated data listings) in the CCE LTER catalog are published in the LTER network archive.

Data Products, Curation and Data-Use Policy

All raw data will be made freely available by the time of publication or the end of the funding period, consistent with NSF policy. Because we will be collecting samples and data outside standard CalCOFI and CCE LTER operations, we will augment our existing data serving capabilities to include products from the proposed reseasrch. Data originally recorded on physical paper datasheets will be transferred each day into spreadsheets using non-proprietary software (e.g. open office platforms stored as ASCI files, .txt or .csv formats). Data will immediately be checked for outliers in the R statistical program, and any outliers will be checked against the paper copies for transcription errors. Paper copies will be kept on file for at least 10 years. We will also maintain both digital and hard copy versions of CSSI and gut content data generated from laboratory analyses. All laboratory and field computers used for data entry and maintenance are subscribed to the Code42 for Enterprise dynamic cloud back up service. Once project data are ready for publication (cleaned, validated, supplied with Ecological Metadata Standard (EML) metadata), they will be publicly served via CalCOFI's Website, though the SWFSC ERDDAP server and BCO-DMO. Prior to the end of the funding period data will be made available by request with the stipulation that if the data are used in publication then the researchers that collected the data need to be informed of the planned use and be offered authorship as appropriate.

Analysis

Data analysis of field collected data will be scripted and extensively annotated in the R statistical program and the scripts will be version controlled in a Git repository, along with the datafiles required for execution. When data and analysis are associated with a publication, the raw data and associated analysis R scripts will be archived in a source such as Ecological Archives or Dryad.

Physical Sample Preservation

Physical samples collected as part of the proposed research that undergo non-destructive analysis will be submitted to CalCOFI's long-term sample collections at both SWFSC (ichthyoplankton) and SIO (invertebrates) within 1 year of sample analysis.