

Data Management Plan

The proposed work will generate cruise data on water column characteristics, zooplankton distribution, species composition, and abundance, and fish distribution, species composition, abundance, and diet composition. Acoustic data collected during cruises will be used to generate density estimates for each site and sample date, as well as to map biomass density distributions. Model “data” will consist of Bayesian posterior priors for diet compositions and outputs of bioenergetic models.

Prior to developing our own databases, we will consult with data managers at the Biological and Chemical Oceanography Data Management Office (BCO-DMO) to learn how to configure data and metadata so that it is compatible with existing databases. Key aspects of our data formatting for archiving will include (1) following standard formats (2) geo-referencing and (3) data attribution. To the extent possible, metadata for all data sets will be maintained according to data standards (i.e. Federal Geographic Data Committee; <http://www.fgdc.gov/>; NBII Biological Data Profile, The Dublin Core Metadata Element Set, and the Directory Interchange Format), and will include the latest version of Ecological Metadata Language (EML). The intent is to compile metadata-based systems that foster flexible structures for data acquisition, data transformation and analysis. Ecoinformatics tools that structure workflow and permit re-assessment of analyses (e.g., Kempler, <https://kepler-project.org/>) will be employed to the extent possible.

Cruise data

During sampling surveys, we will collect four frequency acoustic backscatter data and water column characteristics data (CTD + dissolved oxygen). Raw acoustic data will be processed to produce density estimates and distribution (density by depth) maps by survey transect. These data will be added to a database (MS Access or MySQL) for our own analysis. Metadata will be compiled in a similar database framework. Zooplankton samples collected during cruises will be processed in the laboratory, and when combined with acoustic data will provide information on zooplankton species composition and taxa-specific densities. These data will be aggregated by site and sample date and added to the project database. Fish data (abundance, body condition, size composition, age composition, diet composition) will be also be aggregated at the site / sample date level and encoded in the project database. All data will be provided to BCO-DMO following completion of our work and publication of manuscripts. A separate database containing stable-isotope ratios will be created. This database will be made available to researchers that request it, and will be stored on a networked hard-drive (terastation) or a cloud computer service (e.g. Amazon ECS).

Model and Statistical Analysis

We will use widely available open-source software (R) or other programs that are commonly used (Matlab) for modeling and statistical analysis. Both permit direct ODBC connectivity to databases so analyses can be run directly from the database. Ecoinformatics tools can help structure workflow so that others may re-run analyses in

exactly the same manner. Publications will detail specifics of models, and we will make available all computer code to researchers that request it.

We will endeavor to adhere to the data management policies of the NSF Division of Ocean Sciences, and to practice data sharing consistent with University of Washington policies governing intellectual property, copyright and the dissemination of research products.