Data Management Plan

Types of data
We will be collecting high-resolution plankton imagery, biological samples, and environmental data via the ISIIS, MOCNESS, and shipboard sensors. All of these data will be collected simultaneously, along with ship's position and time, and will be recorded, stored, and backed-up in digital formats on board the ship. Data storage will take two forms: 1) all cruise data will be transferred first to an Amazon Web Services Snowball (described in Project Description, p. 11) to be shipped immediately after the cruise conclusion for image processing; 2) data backups will leave the ship for storage in laboratory and off-site locations (e.g., OSU CGRB – see below) The ISIIS imagery data are taken at a very high scan rate. The scanned images are initially broken up into an Audio Video Interleave (.avi file extension) containing 430 frames and 1137 frames on the large and small cameras, respectively. The .avi format is used to stack the image frames into a single file to ease data transfer (i.e. avoid read/write issues due to large number of files). The raw files (.tiff image files de-stacked from .avi's) are isolated, treated for background subtraction, thresholded, and analyzed to separate regions of interest (ROIs). These ROIs are then used for classification/identification into targeted categories (e.g., 'fish_clupeid', 'copepod_calanoida', 'copepod_cyclopoida_oithona', 'hydromedusae_liriope', 'protist_radiolarian', 'appendicularian'). Roughly 128 TB of raw image data will be generated for this project.

All plankton samples will be stored in the HMSC Plankton Ecology Laboratory until they are processed and zooplankton and larval fishes sorted and identified. Sorted and identified larval fishes will be stored in vials in the laboratory until selected for otolith microstructure and/or gut contents analyses. Otolith microstructure data will be maintained in the HMSC Plankton Ecology lab in a separate database (with duplicate backup protocols as described above). All samples will be linked among databases via unique station and sample identification numbers. Once gut contents and otolith data are obtained, physical samples will be available for use by others via direct communication with PIs. Resultant data from gut contents, otolith microstructure, and stable isotope analyses will be accompanied by well-annotated metadata in laboratory notebooks, excel spreadsheets and word documents to maximize the usable lifespan of the data. Data originally recorded on physical paper datasheets or laboratory notebooks will be scanned daily. All raw and processed data will be stored on external hard drives in both labs (redundant copies) as well as offsite at the Center for Genomic Research and Bioinformatics.

Data and Metadata Standards
We have previously worked with the Center for Genomic Research and Bioinformatics (CGRB) at Oregon State University to develop an image database platform/structure. This system captures metadata for all original and derivative images processed by the system. These metadata include geolocation as well as parameters such as conductivity, temperature, depth, dissolved oxygen, photosynthetically active radiation (PAR), and fluorometry. These data will reside in a RDBMS (relational database management system) and will be available through the ISIIS user interface and other system features. Where possible and applicable, these metadata will also be made compliant with existing metadata standards, such as CF-netCDF or the North American Profile of ISO 19115:2003 for geographic information, to make the data more easily

consumable by the broader scientific community.

Policies for access and sharing and provisions for appropriate protection/privacy

The images derived from the live camera feed will be stored in a standard Unix file system format with a RDBMS used for file indexing, location, and user access and control (e.g., PostgreSQL; with a highly interactive database schema). Presentation of the data will be provided through a standard web based CMS (Content Management System). We will also work with BCO-DMO both with respect to uploading the data and to coordinate access to the broader scientific community. The data will be made public within a 2-yr time period following initial collection. As these images do not contain any human information, there are no ethical or privacy concerns.

OSU provisions for re-use, re-distribution

We are planning for permanent sharing and reuse of these data. It is likely that many researchers within both the marine ecology and biological oceanography fields could utilize these data. Additionally, these data and our data-sharing system may have applications in other areas of marine science, education, and conservation management. We anticipate that the system we establish for sharing and archiving will also serve for reuse/ re-distribution. A web form will be made available to interested researchers. This form will be published with a persistent URL so that the form will be available for publication or other method of distribution as well as for future reference (Note: this is all in addition to BCO-DMO held data).

Plans for archiving and Preservation of access

This project will generate some of the first ever images of plankton for the northern California Current. We intend to use these data to build a *Global In Situ Plankton Image Library* database for public access. This repository will be housed at the OSU CGRB in Corvallis, Oregon. The repository will be built using open-source tools so that ongoing maintenance and support will not be tied to a proprietary system. The facility is earthquake proof with onsite generators and redundant network connections both within the facility and to the Internet. Original source data will be kept intact. Derived data will not be backed-up as we will re-derive from the original source as needed. Metadata will be kept and backed up with project data regularly within the system. An index of metadata will be created and maintained.

All of our oceanographic/environmental data will be compiled into correct formats (per National Oceanographic Data Center, NODC) and sent to the NODC within one year of the completion of the cruises. Archiving of zoo- and ichthyoplankton data will necessitate sorting and identification of samples, which will require more time. Once complete and quality-controlled, these data will also be archived in national databases (BCO-DMO) where the data will be freely available to the public. Archived biological samples will reside in the OSU Plankton Ecology Laboratory until transfer to a permanent storage/curation facility such OSU Ichthyology Collection in Corvallis, Oregon.