

## **Data Management Plan**

### **I. Types of data**

This project will collect high-resolution plankton imagery, ichthyoplankton, 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. Digital copies of all data will leave the ship for storage in laboratory and off-site locations (e.g. UM CCS – see below)

The ISIIS imagery data are taken at a very high scan rate and then isolated and analyzed to separate regions of interest (ROIs). These ROIs are then broken into separate images for classification/identification into targeted categories (e.g. plankton taxa). Roughly 150 TB of image data will be generated for this project. We will use industry standard image formats, either lossy or lossless compression depending on the contents of images.

All ichthyoplankton samples will be stored in the Cowen laboratory until they are sorted and larval fish identified. Sorted and identified samples will be stored in vials in the Cowen laboratory until selected for otolith microstructure and/or gut contents analyses. Selected samples will be transferred to the Sponaugle laboratory for otolith extraction and otolith microstructure interpretation. Otolith microstructure data will be maintained in the Sponaugle lab in a separate database (with back up 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.

### **II. Data and Metadata Standards**

Under separate funding through NOAA NRDA (related to the Deep Water Horizon), the PI has worked with the Univ of Miami's Center for Computational Science (CCS) 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 reside in a RDBMS (relational database management system) and are available through the ISIIS user interface and other system features. As part of the ISIIS-NRDA project, where possible and applicable, these metadata are also 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.

### **III. 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*). Presentation of the data will be provided through a standard web based CMS (Content Management System). The PI will approve access to the data

until such time as the data are to be presented to the public (see below). The following process will handle access to the broader scientific community: 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. Upon approval, the interested researcher will be given read-only access to the section of the dataset that they have been approved for. In the near term, the PIs retain the right to use the data before access is given to the larger scientific community or the public. The data will be made public within a two-year time period following completion of the funded project. As these images do not contain any human information, there are no ethical or privacy concerns. All intellectual property rights are reserved as per standard University of Miami policy.

#### **IV. Policies and provisions for re-use, re-distribution**

We are planning for permanent sharing and re-use of these data. It is likely that many researchers within both the marine biology and biological oceanography fields could utilize these data. Additionally, these data and system may have applications in other areas of marine science and conservation management. We anticipate that the system we establish for sharing and archiving will also serve for re-use/re-distribution.

#### **V. Plans for archiving and Preservation of access**

We intend to further develop and implement a large-scale data repository of marine silhouette images. This repository will be housed at the University of Miami's Center for Computational Science. The repository will be built using open-source tools so that ongoing maintenance and support is not tied to a proprietary system. The repository will be physically located at the University of Miami's co-location facility, Terremark's NAP (Network Access Point) of the Americas located in downtown Miami. The facility is Category-5 hurricane 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 source as needed. The RDBMS and other components are backed up daily. Meta-data will be kept and backed with normal data within the system. An index of meta-data 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 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 samples will reside in the PI's laboratory until transfer to a permanent storage/curating facility such as the SEAMAP Plankton Storage facility.