Data Management Plan

All oceanographic context data from instruments and collected by bottle samples during the OFP and BATS cruises are subject to BATS and OFP data storage policy and procedures. These data repositories will be linked to data collected in this project through the Biological and Chemical Oceanography Data Management Office (BCO-DMO). Where OFP and BATS are relevant to the classification of images, and where OFP and BATS data are used in models and statistical analyses to derive flux estimates, these data will be joined with the image data in the repository of The Biological and Chemical Oceanography Data Management Office (BCO-DMO) at Woods Hole (bco-dmo.org) under the name of the principal investigator (Bochdansky)..

Data description and format

Metadata for each data set will include the cruise number, date, time (UTC), longitude, latitude, depth, a detailed methods description, and will provide links to CTD data sets, and data resources of collaborators in the same project.

Specific data type (file format):

1. Raw images from CTD profiles and the sediment trap camera will be stored unaltered (bitmap or uncompressed tiff formats) on the Research Mass Storage server at the High Performance Computer Center at ODU.

2. Particle characteristics from the sediment trap camera and the CTD camera (e.g., number per volume of water, length, width, image area of each particle in pixels, perimeter, porosity, aspect ratios ratio, roundness, fractal dimensions, dimensions after conversion to metric units) will be combined with all metadata (e.g., longitude, latitude, Julian day, UTC, and depth (comma delimited text) for each particle. Data will be stored in comma-delimited format.

3. Classified images particles will be stored in bitmap or lossless tiff compression formats in folders named with the classification name at the highest resolution even if classes of particles are later combined for analysis.

Intended repositories:

(1) The High Performance Computing facility at Old Dominion University currently provides 1.2 Petabytes of mass storage for researchers. All data derived from this project (raw images, classified images, processed data, CNN models) will be deposited on the Research Mass Storage server at the HPC. The mass storage data resides on an Isilon storage system comprised of 11 storage nodes connected with an Infiniband backend. Initially, this system provides over 500 Tb of available storage for each ODU researcher to be used with up to 40 Gb of network bandwidth for transferring data to the clusters. Usage can be expanded beyond the 500 Tb upon request. All data including all raw images will be stored on this server which is backed up regularly.

(2) The Biological and Chemical Oceanography Data Management Office (BCO-DMO) at Woods Hole (bco-dmo.org) will serve as a repository for all data except for raw image files, including context data and metadata from the OFP and BATS cruises used in image analysis and deep learning. Data from Niskin bottle collections, microscopic examination and from catalyzed reporter deposition fluorescence in situ hybridization will be uploaded to BCO-DMO.

Timeline for data release:

12 months after fieldwork for oceanographic context data and raw images. Up to 36 months for classified and processed image data, and microscope observations.

Policies for access and sharing

Data will be made freely accessible to the public according to NSF and BCO-DMO procedures and guidelines. All original image files will be made available through the HPC facility at ODU upon request (registration and login required).