Data Management Plan

Particle numbers and characteristics

Data description and format

Metadata for each data set will include the cruise number, depth, 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-attached cameras will be stored unaltered (bitmap format) on the Research Mass Storage server at the High Performance Computer Center at ODU.

2. Raw data of particle characteristics will be stored for each particle in the comma-delimited (comma separated values, csv) format including all metadata.

3. Particle numbers and summary characteristics in bins of 1 m (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 the comma-delimited (comma separated values, csv) format.

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 (including experimental and laboratory data) except for raw image files.

(3) The PANGAEA Data Publisher for Earth & Environmental Science hosted by the Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research (AWI) and the Center for Marine Environmental Sciences, University of Bremen (MARUM) (pangaea.de) will be used as a repository for all expedition data on the RV Pelagia except for the raw image data.

ATP and other variables collected during field operations and laboratory experiments

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):

ATP values will be combined with longitude, latitude, depth, date, UTC and stored in a comma-delimited (comma separated values, csv) format.

Intended repositories:

ATP files will be stored at BCO-DMO and PANGAEA (see above).

Timeline for data release:

12 months after fieldwork for oceanographic context data and raw images. Up to 24 months for classified and processed image data.

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 is required).