

Data management plan

1. Types of Data

Data collected from the project will include observations from both field and laboratory experiments. PI Nayak will lead the overall data curation and storage efforts. All data will be collected, analyzed, and stored by the PIs at Florida Atlantic University (FAU), Woods Hole Oceanographic Institution (WHOI) and the University of Southern Maine (USM). The PIs will register this project with the Biological and Chemical Oceanography Data Management Office (BCO-DMO) by submitting required information. All metadata and complete raw and derived datasets will be made publicly accessibly within two years of collection or earlier on the BCO-DMO website, as per NSF guidelines.

1.1 Field Data

The primary field data products resulting from this project would be: (a) raw images from the stereophotogrammetry and holographic imaging systems; and (b) flow measurements obtained using a Nortek Aquadopp HR profiler. Stereophotogrammetry will provide detailed information about mysid distributions and swarm dynamics, while the holographic system will record data regarding the local particle/plankton community composition. The Nortek Aquadopp HR profiler will provide information about the small-scale velocity distributions. Other ancillary data would include images from a downward-facing camera providing top views of the mysid swarm, physical measurements from the CTD, as well as single point velocity measurements from the ADV. CTD data will include vertical profiles of temperature, salinity, turbidity, fluorescence, and PAR. In addition to our CTD measurements, we will obtain biological data in the form of zooplankton community abundance and composition.

1.2 Laboratory Data

The primary laboratory data will consist of the following (a) high-speed 2-D PIV data of the feeding and swimming currents generated by free-swimming *N. americana*; (b) the lab flume experiment data including image data from the stereophotogrammetry system of the swarming behavior of mysids; (c) 2-D PIV measurement data of the bottom boundary layer flow in the flume, and (d) data from feeding experiments such as the number of individual prey species consumed under various treatment conditions.

2. Data Formats

The raw data collected as part of this project will be in the form of a time series of images from the different imaging systems or from the different standard instruments (ADV, Aquadopp HR profiler and SeaBird CTD). All images will be stored as TIFF, JPEG, or bitmap files. Files from commercially sourced instruments will be converted to ASCII format text files, spreadsheets, or other standard formats. Most raw and processed files will be stored in formats that are widely used in word processing, digital imaging, and database management. For biological data from field surveys such as species abundance and composition, data will first be recorded on printed data sheets and then entered into standard electronic spreadsheets. Biological data from feeding assays will also be entered into both printed and electronic spreadsheets. Scanned copies of all printed datasheets along with electronic spreadsheets will be made available.

Expected data product #1

Data type: Observational, experimental; **Responsible Investigator:** Aditya Nayak, Houshuo Jiang.

Product description: Stereophotogrammetry data from field and laboratory experiments and holographic imaging data from field experiments.

Preservation plan: The image data being collected from both these systems will be memory intensive; thus, the PIs have budgeted for memory storage space accordingly. We anticipate several TB of raw data and thus adequate planning to store the data securely is vital. In the field, as soon as data is collected, the

team will immediately start creating a copy of all files. Once we are back in the office, a third copy of all the data will be stored securely on the cloud. FAU-HBOI also has a disaster proof bunker (as the area is prone to hurricanes), where a copy of all data collected by the PIs during field campaigns will be archived. We will create videos of the time series of images which will be compressed into web friendly format for public access. Compressed raw data files and derived data products (such as spatial statistics, mysid distribution data, etc.), will be archived at the BCO-DMO.

Timeline for data release: All data will be uploaded on the BCO-DMO website periodically and as per NSF guidelines. Results will be disseminated via published manuscripts, publicly available final reports to NSF, and data archiving in appropriate databases.

Expected data product #2

Data type: Observational, experimental; **Responsible Investigator:** Houshuo Jiang.

Product description: Velocity data from PIV (Lab measurements)

Preservation plan: The image data being collected from the PIV system will be memory intensive and handled in a similar fashion to the other imaging systems described earlier. PIV data analysis will entail following standard correlation and post-processing algorithms to produce a time series of 2-D velocity distributions. Further derived data would include mean velocity profiles, visualization of coherent structures, Reynold stress distributions and other turbulence statistics. All compressed raw images and derived data products will be archived at the BCO-DMO website.

Timeline for data release: All PIV data will be made available on the BCO-DMO website periodically and as per NSF guidelines. This data will form the primary basis of manuscripts and other research products which will be disseminated through national and international scientific meetings, seminars, and symposia.

Expected data product #3

Data type: Observational; **Responsible Investigator:** Aditya Nayak.

Product description: Field velocity data (Aquadopp and ADV), CTD data, and optical data.

Preservation plan: Acoustic data from the ADV and the Aquadopp HR profiler would generate velocity distributions and turbulence statistics. The CTD data will be backed up in triplicate as described previously. Data from the ac-s will be extracted and stored in text files. All raw and derived data products from these instruments will be archived at the BCO-DMO website.

Timeline for data release: Raw data will be uploaded to the BCO-DMO database periodically, within three months of a field data collection effort. Processed data will be uploaded as completed, no later than the end of the final project reporting period.

Expected data product #4

Data type: Observational; **Responsible Investigator:** Rachel Lasley-Rasher.

Product description: Biological data (zooplankton abundance and composition)

Preservation plan: We will preserve and archive an aliquot from each sample in case we need to refer back to our samples. Once these samples are visually inspected, abundance and composition data will be recorded into printed spreadsheets and electronic spreadsheets. Scanned copies of printed datasheets along with files of electronic spreadsheets will be made available on BCO-DMO website.

Timeline for data release: One year within collection.

Expected data product #5

Data type: Experimental; **Responsible Investigator:** Rachel Lasley-Rasher.

Product description: Biological data from mysid feeding assays.

Preservation plan: Number of prey consumed in various treatments will be recorded on printed spreadsheets and electronic spreadsheets. Scanned copies of printed datasheets along with electronic spreadsheets will be made available on BCO-DMO website.

Timeline for data release: One year within collection.