

## **Data Management Plan**

### **Types of Data Products:**

This project will produce large data sets, on the order tens of TB, since each experiment will produce tens of thousands of images each roughly ~3 MP in size. The raw images will be stored in TIFF format. Image data will be processed to obtain copepod tracks, particle properties, and other auxiliary metrics. Processed image data will be converted to MATLAB format for use, storage, and sharing with outside parties. MATLAB is commonly used within the oceanographic community, and open-source software (Octave) is available that can read MATLAB files. MATLAB code will also be developed as part of this project. Other auxiliary data will be kept in Excel format.

### **Data Storage and Preservation:**

All raw and processed data, including raw images and processed MATLAB files, will be stored on a 32 TB RAID (Redundant Array of Independent Disks) currently in possession in PI Prairie's lab (G-Speed Q by G-Technology). MATLAB code will similarly be stored on the RAID drive in addition to a repository on GitHub. A RAID drive is a reliable method for storage of large amounts of data, particularly because of the built-in data redundancy. In addition, the PI will work with members of Information Technology Services at the University of San Diego to set up a password-protected server with local-area-network service, on which the RAID drive will be backed up at least weekly. All data types will be archived in perpetuity.

### **Data Sharing and Access:**

All collected data will be managed with the assistance of the Biological and Chemical Oceanography Data Management Office (BCO-DMO, <http://www.bco-dmo.org>), which was created to serve data online from research projects funded by the Biological and Chemical Oceanography Sections for dissemination and protection. The PI will work with the office on data quality control, metadata formatting, and data exchange.

All processed image data and raw auxiliary data will be made available within 2 years of acquisition. MATLAB code developed as part of this project will also be provided to BCO-DMO, in addition to series of sample images so that the code can be run by any interested outside parties. Given the quantity of data associated with this project, it will not be feasible to share all raw image data online. However, the PI will work closely with BCO-DMO to provide as many images as possible, which will be aided by the compressibility of TIFF files without the loss of much information. The remainder of raw image data will be made freely available upon request if external hard drives are provided by the interested parties.

The dissemination of summarized results from this project will primarily be through peer-reviewed publications. Research results will also be presented by the PI, postdoctoral scholar, and students at 1-2 national and local scientific conferences per year. Lastly, MATLAB code developed as part of this project will be uploaded onto a repository on GitHub and then, following publication of results, will be released on Zenodo, a data archiving tool which assigns a digital object identifier (DOI) so that MATLAB code in the PI's GitHub repository can be easily accessed and cited in the academic literature.

The status of data archiving and sharing will be provided by the PI in all NSF Annual Project Reports as stipulated by the Division of Ocean Sciences (OCE) Sample and Data Policy.