Data Management Plan (DolMICROBE)

I. Title and contact Information

Project Title: The significance of doliolid microbial interactions: Do doliolids fundamentally alter the trophic structure and productivity of sub-tropical continental shelf food webs?

Data Collection Title: DolMICROBE

Contact: *Marc E. Frischer, Professor, University of Georgia, Skidaway Institute of Oceanography*
10 Ocean Science Circle, Savannah, GA 31411
Office: (912) 598-2308, Fax: (912) 598-2310
Email: frischer@uga.edu

Frischer will be responsible for the management of all project data and will ensure that the project conforms to NSF policy on the dissemination and sharing of research results. The SkIO IT department will implement and be responsible for maintaining the DolMICROBE dataset, and will assist if necessary, with interfacing with the Biological and Chemical Oceanography Data Management Office (BCO-DMO).

II. Types of data

A. Data description: Three types of data and materials will be generated during the proposed studies and will all be made accessible via BCO-DMO and by direct request. 1) Oceanographic data collected during the repeated and transect cruises on the SAB continental shelf (ADCP current profiles, depth profiles of temperature, salinity, chlorophyll fluorescence, dissolved oxygen, light, turbidity, image based-zooplankton abundances and classifications (including doliolid life stages), and particle abundance and sizes). 2) Biological and chemical data including: sized fractionated POC concentration and isotopic composition, extracted chlorophyll a, bacterial productivity, light and dark respiration rates (net heterotrophy), photosynthetic activity, and microscopy determined abundance of doliolids and copepods. Doliolid and ocean associated microbiomes (16S rRNA-based) and ocean metagenomes and metatranscriptomes will also be characterized. All sequence data generated will be deposited in GenBank and Dryad. Isotopic composition of doliolids and other representatives of the zooplankton communities will also be collected. 3) Experimental data including growth, reproduction, and assimilation of C and N from experimental diets will also be collected, archived, and shared. All data and metadata will initially be recorded in logbooks and stored as simple Excel spreadsheets. Data from the mDPI (mini Deep-focus Plankton Imager) system will consist of raw data files, including .csv for physical oceanographic data and .tiff image "stacks" for raw data collected with the instrument. In addition, we will provide all code used to process and visualize the data obtained from the mDPI. The machine learning training libraries and manually validated images will be saved as individual .tiff files. The physical environment of each 'particle' or individual zooplankter will be in .csv format and include the location of the bounding box of each individual along with the oceanographic properties associated with the surrounding fine-scale environment. With the exception of the image data, which we estimate will reach 10-40 TB, we anticipate that other data will not exceed 0.5 TB over the course of the study.

B. Existing and ancillary data: Relevant data associated with our previous studies are available at the BCO-DMO (https://www.bco-dmo.org/dataset/692279). No additional data, software, or special tools will be necessary to utilize the DolMICROBE project data sets.

III. Data and metadata formats, standards, and organization

A. Formats: Data will be generated from outputs of various analytical instruments or collected by specialized software. However, all data (or summaries of it in the case of mDPI-derived zooplankton community data and bacterial community sequence data) will be exported to common Excel and .csv spreadsheet formats and recorded as hard copies in log books. Data submitted to BCO-DMO will be

easily converted to ASCII format which is the simplest possible format compatible with the tabular format of the data generated. Image segments and machine learning training libraries will be provided as .tiff images in a .zip folder, with their associated timestamps, so a user can determine where the images were taken. No special tools are required to read these formats.

B. Metadata: We will use the existing BCO-DMO metadata forms to prepare the metadata related to the DolMICROBE datasets. Metadata will be used to provide an overview of both the data and collection/processing methods. The aggregate data sets will be accompanied with an overall project description. To our knowledge there are no standard vocabularies, keywords, or conventions other than common names to describe the oceanographic, chemical, and biological data we will collect. If new vocabulary or conventions arise we will contact BCO-DMO support staff to integrate them.

C/D. Data organization/Data quality: The organization, collection, and quality control of data collected or generated in the field will be the initial responsibility of each PI. Frequent project meetings will facilitate the organization and updating of all data sets into a single set of data files organized by the metadata scheme before transfer to BCO-DMO. We will follow the best practices for data management as specified by NSF and our Institutions. Frischer and Brandes have worked together and are familiar with each other's data formats; data standardization has emerged during the process of data dissemination (reports, meetings, publications). Greer is a new faculty member at UGA-SkIO but has been/is a member of large project consortia that involve extensive data management and sharing plans (e.g., GRIID-C). All data files will be placed on a secure (OneDrive) with access by password only. All authorized project participants will have access privileges.

IV. Data access and sharing

Once the data have been quality controlled, we will strive to make all data publicly available. In general, we do not intend to impose any data embargos, with the exception of student generated data that will be used for the completion of student theses and publications. We do not expect that the data we will generate will require any exceptional arrangements due to questions of ethical restrictions or release of indigenous knowledge. The DolMICROBE project will upload data files and collection-level metadata to BCO-DMO as soon as they have been quality controlled and processed.

V. Data reuse

Due to the lack of similar data, we expect the data generated by DolMICROBE to be of interest and use to a wide range of Biological and Chemical Oceanographers with interests in continental shelf processes. Although it is difficult to anticipate the use of DolMICROBE data beyond the research community, the data may also be useful to managers of biological resources in the region. If so, we would assist in interpreting our data upon request. All DolMICROBE data will be described in accordance with current BCO-DMO standards. We will work closely with BCO-DMO curators to ensure accuracy and completeness. BCO-DMO will provide a recommended formal citation for the DolMICROBE data set, including a persistent identifier.

VI. Data preservation

Prior to submission of data to BCO-DMO, all project data will be stored on secure server systems at SkIO. *Please see the description of available IT infrastructure in the description of Skidaway Institute's FACILITIES. SkIO servers are automatically backed up daily to offsite cloud-based storage, and robust disaster recovery policies are in place.* BCO-DMO will ensure that the data are curated in a relevant long-term archive and ensure data will be available after project funding has ended. We will use BCO-DMO tools to create metadata for long-term data preservation.