

Data Management Plan (DolLAYER)

I. Title and contact Information

Project Title: Do fine-scale water column structure and particle aggregations favor gelatinous-dominated food webs in subtropical continental shelf environments?

Data Collection Title: DolLAYER

Contact: Adam Greer, Assistant Professor, University of Georgia, Skidaway Institute of Oceanography (SkIO)
10 Ocean Science Circle, Savannah, GA 31411
Office: (912) 598-2334, Fax: (912) 598-2310
Email: atgreer@uga.edu

Greer, working closely with Co-PIs Frischer, Brandes, and Treible, 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 DolLAYER 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 cruises on the SAB and nGOM continental shelf ecosystems (ADCP current profiles, depth profiles of temperature, salinity, chlorophyll fluorescence, dissolved oxygen, light, turbidity, image based-zooplankton community composition and quantification). 2) Biological and chemical data including: sized fractionated POC concentration and isotopic composition, extracted chlorophyll *a*, photosynthetic activity, microscopy determined abundance of zooplankton, and molecular analysis of prey fields and gut content in the form of 18S rDNA sequence libraries. All sequence data generated will be deposited in GenBank and Dryad. 3) Isotopic composition of zooplankton community, specifically bulk, and amino acid compound specific, ¹³C and ¹⁵N data. All shipboard collected data and metadata will initially be recorded in logbooks and stored as simple Excel spreadsheets. The mDPI data will consist of raw data files, including .csv for oceanographic sensor data and .tiff images for raw data collected with the instrument. The machine learning training libraries and manually validated images will be saved as individual .tiff files. The processed files will be in .csv format and include the location of the bounding box of each Region of Interest (ROI, e.g., segment) merged with the nearest physical data, giving each “particle” its exact location and associated oceanographic environment. With the exception of the image data, which we estimate will reach 20 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> and [/641283](https://www.bco-dmo.org/dataset/641283)). Other imagery data that have been collected in the nGOM can be found at <https://data.gulfresearchinitiative.org>, and appropriate doi's are provided in the metadata. No additional data, software, or special tools will be necessary to utilize the DolLAYER 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) 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 and with most data formats. Image segments and machine learning training libraries will be provided as .tiffs 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 DoLLAYER datasets. Metadata will be used to provide an overview of both the data and collection 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, Brandes, and Greer 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 and Treible have worked on large project consortia that involve extensive data management and sharing plans. 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 collected and quality controlled, we will ensure all data are publicly available within two years. 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 DoLLAYER 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 DoLLAYER 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 DoLLAYER 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 DoLLAYER 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 DoLLAYER 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 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.