Data Management Plan

1. Data types and storage

Data types collected during the proposed research will include: cruise event log, environmental ship underway data, environmental sensor data (i.e. CTD plus auxiliary sensors for chlorophyll, turbidity, dissolved oxygen, etc.), environmental inorganic nutrient concentrations, phyto-, micro-, mesozoo-, and ichthyoplankton biomass and assemblage composition from both current and historic datasets, micro- and mesozooplankton grazing rates, and larval fish diet composition and growth rates. CTD data will consist of raw (.con, .hdr, .hex, .bl) and processed file types (.cnv, .asc, .btl). Cruise scientific event log will include event numbers, start/end dates, times & locations of instrument deployments. Log information will be recorded on paper log sheets (PDF file). Routine underway data (including meteorological, sea surface temperature, salinity, fluorescence, ADCP) will be saved as .csv and ASCII .txt files. Plankton tow numbers, locations, depths, dates, and times will be recorded by hand on log sheets for each net. Information from log will be scanned (.PDF file) and digitized in a Microsoft Excel file. These metadata will be standard across all sample collections and will be agreed upon during cruise planning teleconferences. Photos of each tow/trawl will be taken on the ship using a digital camera (.jpg). Derived data products from plankton tows (i.e. plankton assemblage data) will consist of taxonomic lists (.csv and .accdb files), well as image libraries (.jpg and .tiff files) generated by the FlowCAM and ZooScan, and molecular data will be stored as .seq, .AB1, .sra, .fasta snd .fastq files). Larval fish data, consisting of taxonomic classification, individual length and dry biomass, diet composition, growth rates, will be saved as Excel file. Otolith images will be saved as .jpg files. Larval fish gut content data will be saved as: PDF files of handwritten log sheets; Excel files, and image files (.jpg). Data synthesis products from analyses will also be available as Excel, PDF, .Rdata, and Matlab files. Appropriate and data-specific standards will be used to assure quality assurance (QA) and quality control (QC), and QA/QC metrics (i.e. calibration data, field blanks, etc.) will be provided as part of the metadata.

2. Standards to be used for data and metadata

Project-level and field/experimental metadata will be created for each project data set using standard formats set by Biological and Chemical Oceanography Data Management Office (BCO-DMO) and uploaded along with the data (see below). Collected data will be maintained in machine-readable formats that are compatible with software that are available in public domain, i.e. ascii and/or text-based. Field observation data will be stored in flat ASCII files, which can be read easily by different software packages. Field data will include date, time, latitude, longitude, cast number, and depth, as appropriate. Quality flags will be assigned according to the ODS IODE Quality Flag scheme (IOC Manuals and Guides, 54, volume 3; http://www.iode.org/mg54_3). Metadata will be prepared in accordance with BCO-DMO conventions (i.e. using the BCO-DMO metadata forms) and will include detailed descriptions of collection and analysis procedures. Image data will be stored in JPEG format and PDF. Project-level and field/experimental metadata will be created for each project data set using standard formats set by BCO-DMO and uploaded along with the data (see below).

3. Policies for data access and sharing

The project investigators will comply with the data management and dissemination policies described in the *NSF Award and Administration Guide* (AAG, Chapter VI.D.4) and the *NSF Division of Ocean Sciences Sample and Data Policy*. Existing data and physical samples from the July 2017 GOMECC-3 cruise (CoPIs Stauffer and Schnezter) will be processed and shared with the scientific community in a manner consistent with the Division of Ocean Sciences Sample and Data Policy and the GOMECC requirement to share data within 2 years of analysis . Project metadata for these existing data sets will be loaded into BCO-DMO within 1 month of the proposed project's start date. PI and CoPI lab websites will provide links to request access to the data and physical plankton samples during and after the project period. Oceanographic data will be made available on project websites within 1 month of collection

following sufficient QA/QC screening. Requests for data access will be accommodated in as timely a manner as possible and in collaboration with other project investigators.

4. Policies and provisions for re-use, re-distribution, and the production of derivatives

Project metadata files, full data sets, derived data products and physical collections will be made publicly accessible within two (2) years of collection, unless otherwise specified. This includes software and derived data products (e.g., model results, output, and workflows). The project investigators will work with BCO-DMO data managers to make project data available online in compliance with the NSF OCE Sample and Data Policy. Immediately after completion of the research cruise, underway data and metadata will be submitted to the Rolling Deck to Repository (R2R) project. Data, samples, and other information collected under this project can be made publically available without restriction once submitted to the public repositories. Ichthyoplankton voucher specimen samples will be kept in the TAMUCC Ichthyology collection. In addition datasets will be submitted to the Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC) database, which is a data management system to store scientific data generated by Gulf of Mexico researchers (https://data.gulfresearchinitiative.org/). Data, samples, and other information collected under this project can be made publically available without restriction once submitted to the public repositories. Historic data: GOMECC data will be provided within 2 years of analysis to an archive as described below. Access to SEAMAP metadata is available at http://seamap.gsmfc.org/historic data, and plankton sampling data and samples are available upon request to the NOAA NMFS Pascagoula/MS and through the SEAMAP Ichthyoplankton Collection at FWC Fish and Wildlife Research Institute (http://myfwc.com/research/saltwater/specimen-collections/sis/seamap/). Publication of manuscripts from the project in peer-reviewed journals offering affordable open access options will be prioritized to facilitate access for the public, scientific, and resource management communities. We have requested funds to support this priority for publication (please see Budget Justification). Data sets archived at BCO-DMO will be assigned digital object identifiers (DOI) to facilitate their broader use and distribution among the scientific community.

5. Data preservation and archiving

Raw and processed CTD data, cruise event logs, and underway ship data will be archived at BCO-DMO and the Rolling Deck to Repository (R2R). Digital project data will be stored following a three-two-one storage approach throughout the duration of the project. This approach consists of maintaining three copies of data, with two copies stored on different storage media (e.g. hard drive, DVD), and one copy kept in a separate physical location. Project personnel will store working files on a shared cloud-based system (e.g. Dropbox). This will ensure duplicate copies are maintained on separate media as cloud-based storage services save files on both the user's local disk and a remote server. A third copy of all project-related data will be archived monthly on external hard drive kept at UL Lafayette. Derived image data will be stored in JPEG format and PDF archived in compressed format at BCO-DMO. Oceanographic, biomass, assemblage, grazing rate, and growth rate data will be deposited in open-access formats (e.g. .csv) for long-term preservation in the BCO-DMO and GRIIDC repositories at the conclusion of the project. This will occur no later than following publication and/or within two years of the end of the project. Sequence data will be deposited to GenBank following recommendations by the Genomic Standard Consortium to improve the quality and quantity of contextual information that is provided with next generation sequence data in public databases. All data mentioned above will be additionally backed up on the main server of the University of Louisiana at Lafayette and on external hard drives connected to lab computers on a regular basis. Phytoplankton (CoPI Stauffer), mesozooplankton (PI Robinson) and ichthyoplankton (CoPI Geist) physical plankton samples will be preserved and archived in investigator labs at their respective institutions (UL-Lafayette and TAMUCC).

PI Robinson, CoPI Stauffer, and CoPI Geist will be responsible for ensuring policies and procedures outlined in this data management plan are complied with for phytoplankton, zooplankton, and ichthyoplankton data and physical samples, respectively.