Data Management Plan - This data management plan is crafted around the importance of openly sharing, archiving, and managing data in a fashion that is responsible, timely, and tractable. The project team will adopt data management and analysis practices that utilize robust, well-engineered data management solutions such as Slack and open data repositories. In addition to properly storing and tracking project data, we will integrate data management efforts alongside collaboration and co-production by the project team, including project related communications. Broad dissemination of training and research products will also be a key component of our project's open data management strategy. All data, biological samples, and analysis results will be made available in public archives as described in the sections below.

Data Types and Sources -This project will create several data types, including data that will be novel and of broad significance to the wider community of scientists working within the southern California Current Ecosystem, and further afield in global deep-sea ecosystems. We will maximize project and study reproducibility by archiving and making publicly available both raw data and data products to the scientific community. During the course of the project period, we will generate the following data types:

- **Environmental and Collection Data from MOCNESS tows -** CTD-profiles will be collected during tows, including PAR and oxygen. Sampling depths for MOCNESS nets across different tows, along with dates and times will also be compiled. These data will be presented in standard formats.
- **Micronekton Community Composition Data** - These data will be generated by the graduate and undergraduate assistants under the supervision of the PI and with help from expert taxonomic colleagues at SIO and elsewhere. Data will include taxonomic identifications and accompanying numbers (counts) and biomass (wet weights) from across sampling stations and discrete depths.
- **Size-based Micronekton Community Data** - Individual micronekton specimen lengths and weights will be generated for a diversity of micronekton taxa, particularly for biomass dominant groups.
- **Biochemical Composition Data** - Moisture content, C:N ratios, and elemental C and N compositions will be generated for specific micronekton taxa using an Elemental Analyzer and lyophilizer. These data will be stored in .csv files with accompanying "ReadMe" documents.
- **Bulk Stable Carbon and Nitrogen Isotope Compositions (bulk SIA) - δ^{15}N and δ^{13}C values (per mille) for a diversity of micronekton taxa will be generated by the UC Santa Cruz isotope facility.
- **Amino-Acid Compound-Specific Nitrogen Isotope Compositions (AA-CSIA) - δ^{15}N values (per mille) of individual amino acids will be measured from the bulk tissues of a diversity of micronekton.
- **Computer Code and Model Outputs** - Computer code will be produced strictly in open-source languages such as R and Python. Analysis code will be created to produce isotope mixing pipelines that examine connectivity between depth-discrete micronekton communities. Additionally, code will be developed for statistical analysis and data visualization. Trainees will be taught to develop code in open source R Markdown and Jupyter Notebook files for maximum readability and reproducibility.

Education Materials - The proposed educational activities will produce new curricular materials that may be beneficial to other members of the scientific and educational communities. We plan to set up a course website, hosted within the SIO and UC San Diego domain, where we will make available relevant labs, lecture materials, and photos with blog posts to share lessons and activities from our new courses. This will be done in accordance with University policies that protect student, faculty, and staff's privacy in the dissemination of educational information.

Biological Samples - Biological samples collected for this project will include formalin-preserved micronekton for community composition from MOCNESS tows, selected ethanol-preserved micronekton, and frozen tissue samples for stable isotope analysis. Frozen samples will be destroyed in the process of stable isotope analysis. Formalin and ethanol-preserved samples will be used to verify specimen identifications and overall community composition data. We will work with the faculty Curators of the

various Scripps Institution of Oceanography “Collections” (https://scripps.ucsd.edu/collections) to arrange curation and formal archiving of these samples once we are done with them, making specimen data and their use fully searchable and available to the wider scientific community under established Collection practices and policies. The Scripps Oceanographic Collections at UC San Diego are world-renowned repositories supporting scientific research, educating current and future generations, stimulating curiosity and research, and supplying information to governmental agencies and public policy makers. These collections provide the basis for understanding the ocean's biodiversity, the evolutionary history of life on Earth, and the rates and characteristics of climate change.

Linking Project Communications with Data Management - Communications are commonly neglected when designing data management in large projects. They are critical in providing the context for understanding the provenance, structure, intent, and decisions about data. We will link project communications very tightly with project data using the team messaging app Slack (https://slack.com). Slack is now widely used by academic research labs and comes with multiple inherent advantages related to an open working environment. By default, all team members can see all messages within the team. New team members will gain access to prior correspondences and will benefit from understanding how past decisions were made. All project participants will be invited to view and participate in discussions between team members and collaborators. It also provides tools for voting on decisions, as well as defining and tracking progress toward goals. Slack has integrations for the other tools we will use, which centralizes all notifications. Discussions about data will be explicitly linked to the data themselves. Many industries have already switched from email to Slack, so project trainees will have extensive experience when they finish the program with tools they are likely to encounter in the workplace.

Data Storage and Collaborative Computing - Over the course of the project, data generated from the proposed research cruise and laboratory analyses will be stored in multiple separate places to assure long-term perpetuity. First, physical data sheets will be used across all project tasks and at-sea during field collections. These data sheets will be digitized and files will be stored on a project-specific Team Google Drive, which is provided by UC San Diego at no additional cost to this project. All data files and associated analyses will be backed up on the Choy lab server, which is maintained by UC San Diego and SIO to seamlessly act as a core data backup. Collaborative computing (e.g. pair programming by multiple team members with varying computing backgrounds at a single workstation) will be a standard practice utilized by the project team. This will aid in developing the computing abilities and confidence of diverse learners across career stages. Code reviews will be conducted by project participants, which will minimize analysis errors and build collegiality.

Public Archives for Data Dissemination and Long Term Archival - Permanent analysis files and data generated by trainees (described above in “Data Types and Sources”) will be deposited in public archives. Along with relevant environmental metadata, all community composition, biochemical, and stable isotope data generated by this project will be stored in the publicly-accessible Biological & Chemical Oceanography Data Management Office (BCO-DMO, https://www.bco-dmo.org) once published or within two-years of collection date (whichever is sooner). This practice will facilitate responsible data stewardship, wide dissemination, a higher likelihood of collaboration and knowledge-building, as well as data storage on short and longer time scales. Publicly archived files will be redundantly stored in openly-accessible git repositories associated with each manuscript to ensure data integrity over longer time periods.

Each manuscript published with support from this project will have an associated git or other accessible repository with all code needed to rerun analyses, the data themselves, the text of the manuscript, metadata, and any other files associated with the manuscript. This is already common practice in the PI’s lab. Links to all manuscript data repositories associated with this project will also be provided to BCO-DMO.