Data Management Plan

1. Data/products generated by the project

Our project will focus on data analysis and model building, and we are not proposing to generate new datasets. Existing data sources this project will use include:

- NOAA-Northeast Fisheries Science Center's bottom trawl survey: georeferenced biomass and abundance by species and body size from 1973-present for the Northwest Atlantic Shelf
- NOAA-Northeast Fisheries Science Center's Food Habits Database: georeferenced stomach contents (including prey identity and size) by species and body size from 1973-present for the Northwest Atlantic Shelf
- ROMS model outputs (Kang and Curchitser 2013)—sea surface and bottom temperatures from 1958-2019

From these data sources, we will be compiling processed data sets, including:

- Maps and calculated centroids species' distributions on the Northwest Atlantic Shelf
- Summaries of diet composition for fishes based on size class, season, and year for the Northwest Atlantic Shelf

From the data sources and our processing, we will use models to build projections, including:

- Maps of future species' distributions on the Northwest Atlantic Shelf
- Future diet of select fishes inhabiting the Northwest Atlantic Shelf
- Estimates of energetic quality of future diets.

2. Standards to be used for data and metadata format and content

The PIs and postdoctoral associate (Andrew Allyn, GMRI) will work with BCO-DMO staff to prepare our physical and biological data for serving via BCO-DMO. As applicable, we will conform to standards provided in the Data Management Guidelines Manual. Metadata will provide information on the contents of data files, data source, access information, data organization, and file structure and format, within two years of collection. Data sets to be shared via BCO-DMO will be stored as comma-separated value (CSV) for flat files and as netCDF files for spatial arrays (e.g., distribution model predictions at each grid cell through time) to be useable across many data analysis programs. More detailed documentation of how original source data were processed to create derived datasets will be maintained in R scripts and markdown files and hosted on GitHub.

For analyses and model outputs that fall outside of BCO-DMO's content expectations, the project team will use the GMRI GitHub repository to store code (written in R and Matlab) used to conduct analyses and projections described in the proposal. This will provide version control to the team and facilitate collaboration. When submitting data sets to BCO-DMO, we will also provide links to the GMRI GitHub repository such that the code is discoverable through BCO-DMO.

3. Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property or other rights or requirements

All data will be stored in relevant labs and on the computers of project team members. Files will be routinely backed up via everyone's remote back-up system, Box. Box access is controlled via unique log-on credentials for each user.

The project team will use GMRI's GitHub repository to store the code (written in R and Matlab) to conduct data processing and complete the simulations described in the proposal. Use of GitHub will facilitate collaboration among the team members and will provide version control that will allow us to

clearly extract the code used for the runs that appear in our expected publications. We will provide BCO-DMO with links to GMRI's GitHub repository such that code is discoverable through BCO-DMO.

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

We will make derived environmental and biological datasets available for re-use (with appropriate acknowledgement) and store these, and the original source datasets, on BCO-DMO, within two years of collection.

5. Plans for archiving data, samples, and other research products, and for preservation of access to them

Data sets, R scripts, and Matlab code used for processing and analysis will be archived and backed up on UNH's and GMRI's Box for future use by members of the project team or others who request the data.