Data Management Plan for "Collaborative Research: The consequences of species loss for food web persistence and functioning in the Gulf of Maine rocky intertidal"

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. This data management plan also follows best practices developed by the Center for Research Data and Digital Scholarship (https://www.colorado.edu/crdds/) at University of Colorado Boulder (CU).

Data and Metadata Formats
Data will include those collected from field samples and surveys, computer model output, computer code, educational assessments (surveys), and derived data from existing datasets.

Metadata: We will make use of guidelines for metadata and data developed by and in compliance with BCO-DMO standards, which are designed to promote long-term usability. Across data types in this project, BCO-DMO metadata will capture the responsible persons, units and measurement descriptions, relevant methods and/or protocols, geo-references, inclusive dates and time, as well as detailed descriptions of each variable in the dataset. For geospatial data will use ISO 19100 series standards, specifically ISO 19115 for metadata and ISO 19125 ("simple features") for 2d geometries. For the surveys, data and metadata (code book) for the survey will be stored on the secure, password-protected Qualtrics platform but will be downloaded and converted to ASCII-formatted text files (tab-delimited for data; txt for metadata) for archiving.

File formats: Unless otherwise noted all files (data, source code, documentation) will be stored ASCII (or UTF-8 for non-numeric data) in text formats during the project and for archiving as .csv files.

Derived food web data: Food web data will be gathered using a structured literature review. Each entry in the database will include: species names, their interaction type, site information (latitude/longitude, water depth, wave exposure), study information (type of experiment, study year) and publication information (authors, publication year).

Field surveys and experimental sampling: During field sampling, the data will be initially recorded by hand into waterproof notebooks. After each sampling trip, handwritten data will be entered into electronic spreadsheets. In addition, each month, PDF copies of the data will be archived digitally and with a hard copy. Data collected in the field will include: community data (species names, relative abundances, densities), ecosystem function data (species names and sizes, chlorophyll-a concentration), and feeding data (species names and sizes, names of prey items). Species will be identified using restricted vocabulary from NCBI Taxonomy for Organisms, allowing for dataset integration between the food web data, field measurements, and experiment, and re-use.

Computer code and output: Dee and Barner will be responsible for ensuring code development practices (version control, testing) are followed, including training and mentoring postdoc, GRAs, and undergraduate students. Data compilation and simulations will be conducted using R, and the modeling pipeline will be documented in Github with all R code publicly available. We will develop all code under version control using git (backed and shared on GitHub; password protected), following best practices for code testing, review and documentation. We will include documentation describing the installation and use of code in standard ASCII text formats. For simulation outputs we

will also record metadata that describes the provenance and version of software for each output (the version and program name that produced the file).

Education assessment data: We will implement surveys in the classroom at CU (in Dee's Principles of Ecology course) to assess learning outcomes using a pre- post-assessment design through Qualtrics. We will remove all identifiable information from survey data to protect the confidentiality of survey participants, and all survey records will be anonymized. Where possible we will use Open Science Framework (OSF) to manage all the components of the social data involved in this project – CU is an institutional member of OSF which provides fine-grained access controls for private data as well as the ability to publish to open archives or repositories. These archival files will be kept confidential and access-controlled in OSF, which with CU has an institutional membership. The aggregated, non-confidential data will be shared once any identifiable information has been removed.

Data Storage and Access During the Project
The investigators will store project data (including spreadsheets, computer code and model output, and ecosystem functions sampling data) on CU Boulder and Colby College servers that are backed up by each University's or Department's IT organization. Personal computers in all laboratories are backed up daily to these servers. Dee has access to CU servers and the project's shared, unlimited, Google Drive cloud storage and will be moved to a permanent, public repository post-publication.

Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution
For all datasets, we will adhere to and promote the standards, policies, and provisions for data and metadata submission, access, re-use, distribution, and ownership as prescribed by the BCO-DMO Terms of Use (http://www.bco-dmo.org/terms-use). Immediately upon completion of the project, all species interactions, ecosystem function, and computer model data as well as computer code will be made publicly available through the BCO-DMO data repository. Computer code will also be publicly available on GitHub to allow future versions to be tracked. We will record species information using restricted vocabulary from NCBI Taxonomy for Organisms for consistency and data reuse.

Plans for Archiving
BCO-DMO will ensure that project data are submitted to the appropriate national data archive.

Roles and Responsibilities
The Principal Investigators will take responsibility for the collection, management, and sharing of the research data. Each PI will be responsible for sharing her subset of data among the project participants in a timely fashion, and will coordinate overall data management, sharing, and archiving process. Barner will coordinate submitting field and experimental data to appropriate repositories and providing accession numbers and metadata to BCO-DMO. If Barner leaves the project, Dee will take over these tasks. Dee will be responsible for the educational survey data and will coordinate submitting the computer modeling code and output to BCO-DMO repositories. If Dee leaves the project, Barner will take over these tasks. Should either co-PI leave her current institution, the co-PI will continue to manage and retain the data, including migrating the data from Google Drive storage at CU or Colby to another appropriately secure service if necessary, and depositing it into BCO-DMO data repository.