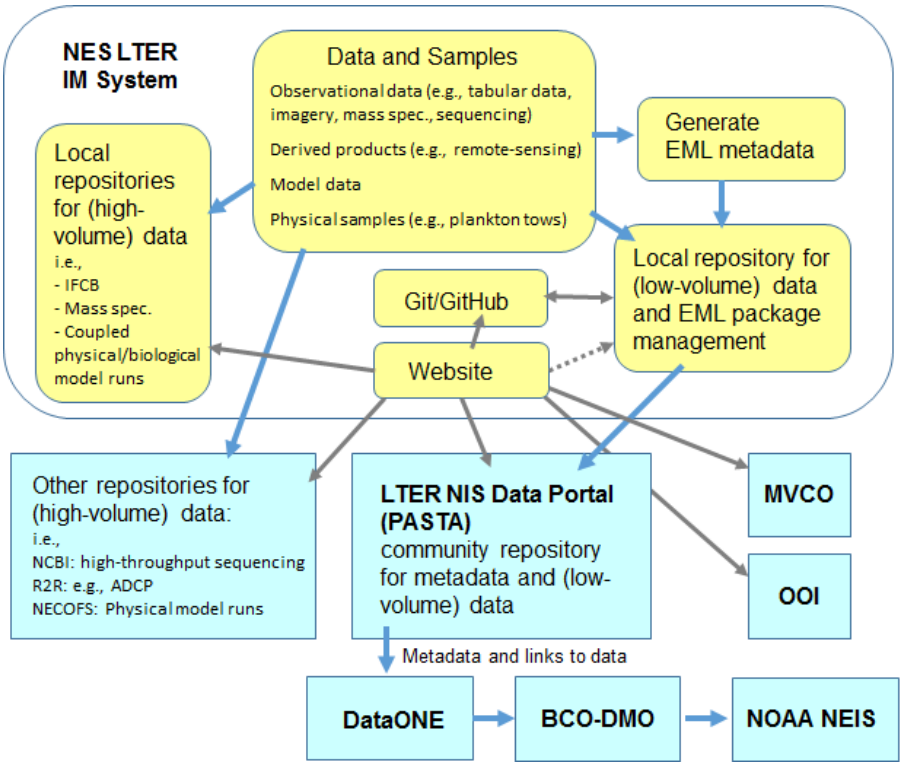
DATA MANAGEMENT PLAN

1. Overview of Northeast U.S. Shelf (NES) LTER Data and Information Management (IM)

The primary goals of NES-LTER IM are to facilitate access to NES-LTER data by scientists, educators, and the public, and to curate those data with metadata to ensure discoverability and usability in the future. For core data sets, we will meet both of these goals by providing data packages inclusive of metadata directly to the LTER Network Information System (NIS). Core data sets will tend to be low-volume and/or low-frequency data that may be derived from high-volume and/or high-frequency raw data. Much of the high-volume and/or high-frequency raw data will also be accessible and discoverable through infrastructure maintained at other community repositories or at WHOI (see diagram).



The NES LTER IM team involves a communication network with an Information Manager, Stace Beaulieu, who will be involved in the entire data lifecycle from planning sample collection to contribution of core data sets and metadata to the LTER NIS. Beaulieu will participate in science planning meetings, coordinate data acquisition with the field research coordinator prior to and following cruises, and train co-PIs and their technicians, postdocs, and students in best practices for contributing data and metadata. To implement the provision of data packages to the LTER NIS and to administer local repositories, Beaulieu will work with WHOI Information Services as described in the Facilities Statement.

2. Types of data produced

2.1. Data. We will produce observational data, derived data products, and model data. Observational data will be obtained in near-real-time from moored underwater instruments, underway and from sampling on research cruises, and post-cruise with laboratory analyses. Observational data and derived products are categorized into the 5 LTER core areas in the Table 2 in the Project Description. A high-volume data example is high-throughput sequencing (HTS) data. A high-frequency data example is imagery from the moored Imaging FlowCytobot (IFCB). The derived core data sets for both these examples will be tabular records of categorized organisms.

2.2. Physical samples will include water samples, filters, plankton net samples, and fish specimens.

3. Data and metadata standards

3.1. Compliance with LTER NIS. Metadata will be provided to the LTER NIS in the Ecological Metadata Language (EML) standard (most recent version). Metadata will be generated by manual entry in a template or tool and may be automated for some data streams. Data provided to the LTER NIS will be in non-proprietary formats, e.g., comma separated values (CSV).

3.2. Compliance with Division of Ocean Sciences Sample and Data Policy. The Biological and Chemical Oceanography Data Management Office (BCO-DMO) will harvest NES-LTER metadata (which will include links to data) through DataONE. Underway data from University-National Oceanographic Laboratory System (UNOLS) cruises will be provided to the Rolling Deck to Repository (R2R). Some high-volume and/or high-frequency data will be provided to other community repositories, for which additional metadata and data standards apply [e.g., HTS data to National Center for Biotechnology Information (NCBI)]. Model data will utilize data and metadata formats familiar to the U.S. Integrated Ocean Observing System (IOOS) community (e.g., NetCDF format). Physical samples will be registered with the System for Earth Sample Registration (SESAR) to obtain International Geo Sample Number (IGSNs) for unique sample identification.

4. Policies for Access and Sharing

LTER-funded data will be made freely and publicly available following guidelines from the LTER Network Data Access Policy for Type I data. Type II data restrictions might apply to products from remote-sensing data if covered under prior licensing and to model data if used by commercial companies. Providers of data access are distinguished as the LTER NIS, other community repositories, or WHOI (see diagram). For example, data from the atmosphere-ocean model and the coupled physical-biological model will be available via THREDDS servers at the Northeast Coastal Ocean Forecast System (NECOFS) and at WHOI, respectively. Each repository may have its own user login for permissions to access during an embargo period not to exceed 2 years after collection. The NES-LTER website will provide links to these and other relevant repositories (dashed gray arrow indicates the possibility for the website to be integrated with a local database, e.g., by adopting a content management system utilized at existing LTER sites).

5. Policies for Re-use, Distribution

The primary policy for NES-LTER data re-use will be the LTER Network's General Data Use Agreement for Type I data. We will recommend citation with Digital Object Identifiers (DOIs) for data sets accessed through the LTER NIS Data Portal, and DOIs can be minted by the MBLWHOI Library for data sets accessed from local repositories. For those data provided to other community repositories, policies of those repositories apply.

6. Plans for Archiving and Preservation

We will facilitate the archiving of core data sets with the National Data Center for OCE-funded data, the NOAA Earth Information System (NEIS). For those data provided to other community repositories, archiving plans of those repositories apply. Local repositories at WHOI's data center will include backup and disaster recovery as described in the Facilities Statement. PIs will archive voucher and type specimens in their labs.

7. Expected milestones and deliverable products from data management

Year	Milestones	Deliverable products
1	Create project website; Adopt template or tool for generating EML metadata; Establish local repository for (low-volume) data and EML package management; Establish GitHub repository for processing scripts for core data sets	Initial EML packages provided to LTER NIS
2	Implement automated upload of EML core data packages to LTER NIS; Establish local repositories for high-volume data, e.g., mass spectrometry raw data; Ensure that our EML metadata can be harvested by BCO-DMO	Additional EML packages provided to LTER NIS; Metadata discoverable in BCO-DMO
3	Establish additional services as needed if more/different data are being acquired; Ensure pathways for ultimate archiving of data in NOAA NEIS	Additional EML packages provided to LTER NIS; Metadata and data harvested by NOAA NEIS