Data Management Plan

This Data Management Plan was prepared in accordance with NSF Data Inventory guidelines for Biological Oceanography. The PI will work with the Biological and Chemical Oceanography Data Management Office (BCO-DMO) to archive and integrate all data from this project and make it available for use. BCO-DMO will assist with archiving ecological data on host density and disease prevalence, and molecular data on genetic diversity.

Types of samples, data, physical collections, and software code
        In addition to BCO-DMO, I will utilize the Knowledge Network for Biocomplexity (KNB: http://knb.ecoinformatics.org/index.jsp), a national network that provides an efficient way to access complex ecological and environmental research data. The KNB network provides access to a database that allows development of metadata specification, Ecological Metadata Language (EML: http://knb.ecoinformatics.org/software/eml/), which was developed by the Ecological Society of America and associated efforts (Michener et al. 1997). EML uses XML documents to organize ecological data into individual modules that describe the project's overall metadata. The proposed research will use EML to facilitate data sharing. To submit data to EML, I will use the Morpho Data Management software (http://knb.ecoinformatics.org/morphoportal.jsp), which will allow me to manage the data to create suitable metadata modules. These databases are easily accessible, as they are based on Java, which works with Windows, Macintosh, and UNIX operating systems, allowing access to multiple operating system users and increasing collaboration efforts with other researchers.
        Ecological Data: Oyster samples will be preserved on ice and stored at -80°C until processing. A subsample of oyster tissue will also be used for genetic and qPCR disease analyses. These samples will generate new ecological data on oyster size, growth and condition. Seagrass samples will be preserved on silica until processing. I will visually quantify lesion area, leaf length, and leaf area and then take a subsample of tissue for genetic and qPCR disease analyses.
        Genetic Data: This research will generate DNA microsatellite data across sites and seasons for oysters and eelgrass. All data tables will contain columns indicating location (GPS coordinates), date, and time of collection, as well as unique site and plot/reef identifiers. All microsatellite data will be archived as GenePop-formatted files (http://genepop.curtin.edu.au/help_input.html). Additional comma-delimited metadata spreadsheets will be generated that link each individual used in experiments to individual plant or oyster identification numbers. We will use microsatellite sequences previously submitted to GenBank (http://www.ncbi.nlm.nih.gov/genbank/), and microsatellite primers and PCR conditions will be published in peer-reviewed journals.
        Human Subjects Data: All survey data will be collected under the auspices of Northeastern University's Institutional Review Board. PI Hughes currently has an active IRB protocols for ongoing surveys of restoration practitioners and natural resource managers (Protocol #16-03-56). Federal Geographic Data Committee (FGDC) compliant metadata will be created and uploaded to the FGDC database (www.fgdc.gov). Northeastern University's Institutional Review Board guidelines will be followed for preparation and archival of data involving human subjects.
        Software Code: Software code used in analysis for publications will be made available via the metadata or other appropriate web repositories (e.g., Dryad) upon publication.

Standards to be used for data and metadata format and content
        I will use standards-based metadata for all measured variables to 1) publicize project datasets through the relevant repositories (e.g. BCO-DMO), 2) classify data according to my data access policy, 3) maintain knowledge of the composition, organization, and quality of the

data, and 4) document detailed data schema to ensure easy and complete understanding of the dataset.

Policies for accessing and sharing data
All data will be uploaded to BCO-DMO within two years of acquirement. When quality control and analysis of the data are not complete within that time frame, the PI will provide the raw data to BCO-DMO and work with the office to temporarily embargo the data until the final data products are ready for final dissemination.

Policies and provisions for re-use, re-distribution, and the production of derivatives
Users will be required to provide their name, affiliation, email address and contact information prior to receiving data, and agree to acknowledge this project and cite the dataset in any publications or derivative projects.

Plans for archiving data, samples, and other research products in a timely manner
Short term: Data sheets will be scanned or photocopied on the day that data are collected, with copies stored in two separate locations. Electronic files will be backed up the day that data are entered, with at least one copy in a remote location.
Long term: All ecological data will be aggregated on a shared cloud server managed by PI Hughes. This server will be automatically linked to BCO-DMO and the KNB. Any remaining oyster and seagrass tissue samples will be stored in the Hughes lab and made available by request. All human subjects data and resulting publications will be archived in Northeastern University's Digital Repository Service (DRS). DRS is a publicly-available digital archive that collects, manages, preserves, and shares the intellectual and historical record of Northeastern University.