

DATA MANAGEMENT PLAN

1. Data Policy Compliance:

NSF requires investigators to provide data within two years of collection. If the project is funded, as a first step we will submit a data inventory form to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) at <http://bcodmo.org/>. This repository's terms of use retain PI's intellectual property rights to data and allow use of data solely for scholarly use by academic and scientific communities, provided the originating Investigator is properly acknowledged and use of data in a presentation, report, thesis or publication requires consent of the originating investigator (PI). Uploads to BCO-DMO will include metadata that includes NSF grant title and number, PI/co-PI contact information, geographic and temporal data, sampling and analytical methodology and pointers to related or linked datasets.

2. Types of data, samples, and other materials to be produced in the course of the project.

Samples of blue crabs and RLV virus will be collected from numerous sampling locations and genetic data from the RLV virus and its host *C. sapidus* will be generated. Oceanographic modeling output will be produced during the course of this project. Experimental results of RLV challenges will be produced.

- A. Biological material archiving. Crabs will be collected and stored either in ethanol or frozen at -80°C in the labs of PI Schott and co-PI Behringer. Some crab samples previously collected in the US and Brazil are stored in the Schott lab and will be retained for this project if it is funded. Subsamples of crabs will be transported to the Plough lab for DNA extraction or extracted in the Schott lab. Extracted RNA and DNA will be stored in -80°C in Schott and Plough labs. Tissue samples and extracted nucleic acids (when available) will be available for sharing with other institutions and researchers. Records of samples (data and metadata) and locations for storage will be in the form of spreadsheets.
- B. Genetic data will be generated from crabs in the form of numerous, short DNA sequence reads (next generation sequencing data), genotype and haplotype data, and a range of population genetic and phylogeographic summary statistics that will be calculated for analysis of population structure and gene flow. Data will be in the form of raw sequence data (Illumina reads), and curated single nucleotide polymorphism (SNP) genotypes in VCF (variant call format) and genepop format, a standard population genetics format. Genetic data from virus will be in the form of complete or partial genome assemblies of short DNA sequence reads (MiSeq or other next generation technology). Assembled genome data will be in text and FASTA format.
- C. Modeling output and data will be generated using the high powered computing facilities at the University of Florida. Output data consists of the modeled trajectories of daily releases of individual crab larvae with positions noted every 12 hours for up to a month. Data will be compressed and stored using NetCDF v4.1. A decade of simulations will be performed and archived from 2004-2015, and the corresponding general ocean circulation model velocity fields from the HyCOM + NCODA global analysis will be stored offline in NetCDF v4.1 format.
- D. Experimental data will be in the form of tabulated results of laboratory based controlled infections. Percent mortality and time to death are examples of experimental data. Results will be contained in spreadsheets and word processing documents.

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

The output data from the Lagrangian modeling will adhere with the latest NetCDF protocols. NetCDF is a machine independent, self-describing, and efficient means of storing and sharing array based data widely used by oceanographers and atmospheric modelers.

4. Methods and policies for providing data access and enabling sharing.

Data generated from the genetic analyses will be stored initially stored between IMET and Horn Point Laboratory, in the Schott and Plough labs. Both labs have continuous backups of information on desktop computers. Data are archived on Over the course of the project and as data analysis proceeds, data (e.g. sample information and location, marker information and sequence data) and eventually results and products will be made publicly available on the PI's (Schott and Plough) website (www.umces.edu) via ESMERLDA, the Environmental Science Metadata-Enhanced Repository for Analysis, Lookup, and Data Archiving available for UMCES researchers. ESMERLDA is a custom n-tier application implemented with a Microsoft SQL Server 2008 R2 back end, that facilitates the, upload, annotation, discovery, and download of scientific data facilitating data sharing and integration between PI's. ESMERALDA has the ability to store a wealth of metadata associated with each project, including descriptions, geospatial and temporal boundaries, units of measure, citations, publications, associated files, and contributors. This metadata is visible via the web interface, and can be downloaded in a number of formats. ESMERALDA is currently housed on a web server at Horn Point Laboratories, and can be used by individuals on the HPL intranet. Preparations are underway to make ESMERALDA visible on the World Wide Web on a trial basis, accessible by users with a username and password.

Genotype data used in publications will be deposited online in DRYAD (<http://datadryad.org/>), and raw, short read sequence data will be deposited in the Sequence read archive <http://www.ncbi.nlm.nih.gov/sra> or other genetic data archiving repositories.

All PIs are committed to an open data access policy and are committed to providing access to such information in a reasonable time frame. During and after the end of funding, essential data will be written up and submitted for publication in journals readily accessible to fisheries or estuarine biology researchers and managers. We will include the progress on the dissemination of data and research products in the final report.

5. Provisions for re-use, re-distribution, and the production of derivatives.

We do not anticipate any restrictions on the use of our data, unless a data source has placed limitations on the re-distribution of those data and materials. We will always review and comply with policies of our data sources to determine the limitations on data re-use and re-distribution. In the event that key personnel leave the project, continued data availability is ensured by the use of ESMERLDA resources at IMET, and by UMCES and UFL policies that stipulate that data generated in these universities remain with the institution.