

Data Management Plan

Woods Hole Center for Oceans and Human Health

PI: John Stegeman, Woods Hole Oceanographic Institution

The Woods Hole Center for Oceans and Human Health is committed to data sharing and technology transfer with other COHH Centers, federal researchers, the public, key stakeholders, and other potential end-users of the results, such as state public health and environmental agencies. We will make all data resulting from this center (projects and cores) available in a format and with documentation/metadata such that data may be used by others. We will adhere to the NSF Division of Ocean Sciences Data and Sample Policy (<https://www.nsf.gov/pubs/2017/nsf17037/nsf17037.jsp>).

1. Products of the research:

Data will include: images and oceanographic data collected from in situ instruments including the Environmental Sample Processor (ESP) and Imaging FlowCytoBot (IFCB), image product data, model output data, standard underway and discrete ship-based and time series measurements, microscope images (fluorescence and electron microscopy), toxin analyses, genetic fingerprinting and DNA/RNA sequencing data, high-speed video data from behavioral assays, web-based mapping tools and associated data, and community engagement and other educational materials.

2. Data format, standards, and accessibility:

The data will be stored in a variety of formats as appropriate to each data type. All data will be collected using best practices and standards as defined by the respective disciplines of the projects and cores. For RNA-seq data, we will follow the best practice guidelines of the ENCODE Consortium (<https://www.encodeproject.org/about/experiment-guidelines/>). The IFCB data require image classifiers to discriminate species and life-cycle stages of species. We will share the classifier data with interested researchers. Other data will be available and accessible through web-based servers or standard software packages that are widely used (e.g. MS Excel, Word). Images and associated metadata will be accessible through html, PNG, JPG, RDF, XML, and other standard formats via web services.

3. Access to data and Data Sharing practices and policies:

We will comply fully with the Public Access Policies of both NSF and NIH.

Data will be accessible through deposition in appropriate databases as recommended by NSF (<https://www.nsf.gov/pubs/2017/nsf17037/nsf17037.jsp>) and NIH (<http://gds.nih.gov/02dr2.html>). All oceanographic data collected during this project will be managed by the Biological and Chemical Oceanography Data Management Office (BCO-DMO; <https://www.bco-dmo.org>). The IFCB data, including image classifiers used to discriminate species and life-cycle stages of species, will be made available in real-time through a WHOI based server (e.g., <http://mellon.who.edu:8888/sp>). RNA-seq data will be deposited into the Gene Expression Omnibus (GEO) database.

All community engagement materials produced by the WHCOHH will be submitted to NSF and NIH through the Partnerships for Environmental Public Health (PEPH) Resource Center, and also shared with program officers at NSF and NIH. Web based mapping tools and associated

data will be available through websites created by the Community Engagement Core (CEC). Community outreach materials and educational activities will be made available online through the Center's websites, and through related CEC websites. Educational materials will be further distributed through networks such as the National Marine Educators Association, Massachusetts Association of Science Teachers, Massachusetts Marine Educators, regional Sea Grant networks, the Woods Hole Science and Technology Educational Partnership, and WHOI's Broader Impacts Group.

Access to some data will be limited to the participating investigators for an initial period of time, but public access to all data and supporting documentation (metadata) will be granted within two years. Data will be released no later than at the time of acceptance for publication of the main findings derived from each data set.

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

Data distributed to the scientific community will be available for reuse and redistribution.

Project 1 maintains a large algal culture collection, including many *Alexandrium* species from the Gulf of Maine and around the world. These have always been openly shared with colleagues, and this policy will continue throughout this project and will include any new isolates of *Alexandrium* or *Pseudo-nitzschia* that are established. In addition, we will provide relevant protocols and published data upon request.

Project 3 will be utilizing several existing transgenic line of zebrafish, but does not anticipate generating any new germ-line transgenic lines. Plasmid constructs that will be generated for transient mosaic expression will be shared with the broader research community by depositing them with Addgene, a nonprofit plasmid repository (<https://www.addgene.org>).

5. Archiving of data:

Each project or core will be responsible for backing up and archiving their data, using procedures appropriate to the type of data being generated. Database deposition is described under (3) above. BCO-DMO will handle submission of oceanographic data to the National Centers for Environmental Information (NEIC; formerly the National Oceanographic Data Center (NODC)) for final archiving at the end of the project. Standard underway ship-based measurements will be archived with the University-National Oceanographic Laboratory System (UNOLS) central data repository at <http://www.rvdata.us/catalog/>, managed by the Rolling Deck to Repository (R2R) project. RNA-sequencing and DNA-sequencing data will be archived in GEO and NCBI databases.