

Data Management Plan

1. The following types of data will be collected and produced in the course of the project

- a) Mortality, morphological measurements (e.g. length, yolk size) and total lipids for Pacific Herring embryos and hatched larvae. These data will then be exported in Excel.
- b) Environmental data including temperature, salinity and pH will be exported to Excel.
- c) Hatched larvae images will be saved as individual images as *.tif format with the following information: magnification, date, sample name, image ID, instrument parameters, scale bar.
- d) LC/MS-MS proteomic data (instrument files: raw spectra, mzML, mzXML), peptide spectral match results including PSM scores, peak areas, and retention time, search parameter files, protein database used to search spectra, and ANOVA and log fold results from treatment comparisons or timepoint analyses.

2. Standards to be used for data and metadata format and content:

- a) LC/MS-MS proteomic data will be provided in standard open mzML format and as convertible raw files. Additionally, all parameter files will be provided as .txt files, and peptide matches and scores will be provided in .csv formats. Final accepted data will be provided in annotated csv files with < 1% false discovery rates, which include 2 tryptic sites, 5 ppm mass tolerance.

3. Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights requirements:

- a) All data sets will be available online within two years of sample collection and analysis to their respective repositories described below. To streamline data accessibility, all chemical (including raw mass spectral files and peak area lists), physiological, microscopy, and laboratory-based culture datasets on various public repositories will be linked to a master project site hosted at Biological and Chemical Oceanography Data Management Office data repository by the PI.
- b) LC/MS-MS proteomic data will be initially preserved on local instruments and servers and available to the public within 6 months of collection on the public Proteome Exchange PRIDE repository.

4. Policies for data use:

- a) Data will be freely available for commercial and non-commercial re-use after publications in peer-reviewed journals with the provision that the NSF grant and citation be properly acknowledged in publications and presentations.
- b) When possible open-access journals will be used to publish work. Publication of data shall occur during the project, if appropriate, or at the end of the project, consistent with normal scientific practices. Research data which documents, supports and validates research findings will be made available after the main findings from the final research data set have been accepted for publication.
- c) Materials generated under the project will be disseminated in accordance with participating institutional and NSF policies. Materials may be transferred to others under the terms of a material transfer agreement.

5. Plans for archiving data/samples:

- a) Data and metadata will be submitted as described above and published, both in print and online as journal articles or supplementary material. In addition, all data files will be retained at least five years after the end of the project, as well as stored at the local home institution.
- b) All proteomic samples will be digested and analyzed using mass spectrometry. Remaining peptides will be stored in a -80°C freezer for up to 7 years at the University of Washington or Western Washington University and available for further analysis if requested by outside labs. All proteomic data generated by PI Nunn will be backed up to tapes are regularly shipped off-site for 3rd party vaulted storage. Tape backups are done using a pair of tape libraries with a combined 22 drives and a local storage capacity of ~6 petabytes.