

## **Data Management Plan**

We are committed to broadly disseminating new data and datasets in a timely fashion, privately among team members (as soon as generated) and publicly with the broader research community (on publication or ~1 year from project completion). We will adopt University of California guidelines for data management <[https://dmptool.org/general\\_guidance](https://dmptool.org/general_guidance)>; data will be managed by Schiebelhut, consistent with NSF requirements (e.g. <<https://www.nsf.gov/pubs/2017/nsf17037/nsf17037.jsp>>). There are no ethical nor privacy issues with the proposed data and no human research subjects are included (i.e. no IRB protocol).

### ***Data management during project***

#### **Specimens**

All collections are made under valid research permits, in accordance with landowner, state, national, and international regulations. All samples are curated with unique identifiers in a fully searchable databased archive in Dawson's lab. The tissue samples are stored in liquid preservative (ethanol, RNAlater) in -20°C freezers that are supplied with emergency power back-up systems in the laboratory. Tissue samples, trackable using the unique identifier, will be shared among partner labs during the project if of use, and with non-project labs via the database (see #5, below) and formal requests.

All field data (e.g. geolocated records, photographs of specimens, additional metadata, etc.) collected by Schiebelhut and Dawson will be archived in original file formats and, if appropriate, exported as tab-delimited text files; the original files are never modified, with all analyses being completed on working copies. These and all other data on personal computers will be backed up in triplicate: to an onsite external hard drive, an offsite hard drive, and the cloud (i.e. Box).

#### **Genomic and transcriptomic data**

All sequence (genomic and transcriptomic) data generated for this project will be backed up on two servers upon being received by the sequencing facility. A third copy will be stored on an external hard drive in Dawson's lab and, when all data come in, all raw sequence files will be uploaded to the NCBI Sequence Read Archive (SRA) for WGS data and Gene Expression Omnibus (GEO) for expression data.

***To ensure data availability for public use and potential secondary uses, there will be no restrictions on sharing, using, or re-using our data. We will:***

1. share sequence and other data with collaborators pre-publication to facilitate their own independent or collaborative efforts to fully describe project research results.
2. upload sequence data to the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) shortly after generation and marked for public release on acceptance of manuscripts or 1 year post-project.
3. upload genetic datasets used in published analyses to Dryad, along with code and conditions for generating the analyzed dataset from the original SRA (see #2), available publicly on acceptance of manuscripts.
4. register the project with the Biological and Chemical Oceanography Data Management Office (BCO-DMO) and curate all sample data, project metadata, and links out to datasets in appropriate genetic archives (e.g. see #s above). We have an established history of depositing data with BCO-DMO before and during publication for other funded projects, and will continue this practice.
5. make relevant entries from Dawson's laboratory sample database publicly available on the lab website as a tab delimited text file amenable to regular expression queries.
6. publish in open access formats to the maximum extent possible, using UC Merced's read-and-publish agreements, e.g. with Wiley and Elsevier, or reduced rates negotiated with other journals (see <[http://osc.universityofcalifornia.edu/alternatives/submit\\_work.html](http://osc.universityofcalifornia.edu/alternatives/submit_work.html)>)