

Data management plan

We follow University of California guidelines for data management <http://www.cdlib.org/services/uc3/datamanagement/>.

Within that framework, each PI is responsible for managing her/his subset of data, with overall management coordinated by Dawson.

There are no ethical and privacy issues with the proposed data and no human research subjects are included as part of this study (i.e. no IRB protocol).

Data management during project: collection & storage

Grosberg and Dawson laboratories' tissue / DNA samples are stored in duplicate in separate -20°C or -80°C freezers that are supplied with emergency power and/or liquid nitrogen back-up systems. We also maintain working archives in -20°C freezers in our laboratories.

Personal computers in all laboratories will be backed up daily using Apple Time Machine to an onsite external hard drive, and weekly to an offsite hard drive. For this project, Dawson has established an account with the San Diego Super Computer's enterprise class Cloud Service (<https://cloud.sdsc.edu/hp/index.php>) for data storage and sharing. Thus, we maintain all sequence data in at least three locations, including one lab computer, one personal computer, and one 'cloud'-based server. All three 'localities' are themselves backed up on a daily-to-weekly basis.

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] upload all sequence data to GenBank for release on first publication, these will be accompanied by all metadata describing collection localities
- [2] deposit environmental datasets and metadata to Dryad (the majority of non-genetic data that we will use in this project are already archived by NOAA, CenCOOS, etc; we will not duplicate any of the raw data in archives, instead focusing on depositing the down-sampled environmental datasets we generate for corresponding genetic analyses),
- [3] explore deposition of our datasets and outline of results on the PISCO data catalogue, online at <http://www.piscoweb.org/data/access-and-applications>,
- [4] make relevant entries from Dawson's lab sample database publicly available online (<http://nscs01.ucmerced.edu/cs/shelflife/login?>),
- [5] make Ph.D. theses available electronically 1 year after filing via the University of California Library,
- [6] publish in open access journals when possible, and
- [7] distribute summary results to California's Ocean Protection Council and other relevant agencies.

Dawson works closely with the Library at University of California, Merced, on a variety of Open Access initiatives, and this will continue during this project, per the University of California's recommendations for data management. If services above are found to be lacking, then we will use the University of California & California Digital Library's UC3Merritt curation services (<http://www.cdlib.org/services/uc3/dmp/index.html>).