

DATA MANAGEMENT PLAN

Types of Data Products

The key data products generated by this project will be monthly-resolved coral geochemistry records for Christmas Island, Symbiodinium ITS2 amplicon sequences and benthic survey data derived from on site surveys and photographic records. Metadata on survey designs will be crucial to interpreting the data, and our management strategy ensures consistent formatting and archiving of linked data products. All coral geochemical data and associated metadata (Xrays, photos, etc) will also comprise product datasets. Ancillary data products from other programs, including oceanographic data, reef habitat data, and meteorological data, will also be important in interpretation and will be linked through open-access data repositories as appropriate and will facilitate interaction with the complimentary AGS/OCE RAPID.

Data and Metadata Standards

Formal Standard Operating Procedures (SOPs), published as peer-reviewed methodologies where applicable, will be developed and made available through participating institutional websites. Quality control will be conducted at each stage of the data acquisition, processing and analyses, including the development of metadata forms detailing the outline of the project, instrumentation used, format of data, QA/QC standards and controls, and funding source amongst other details. Molecular/genetic data and metadata will meet standards of the Genomic Standards Consortium recommendations (MIMS, MIMARKS, and the draft recommendations of MINSEQE). The data and metadata repository available through the Knowledge Network for Biocomplexity (KNB <http://knb.ecoinformatics.org/index.jsp>) an international online data repository. Long-term archival and curation of genetic metadata will occur through GeoSymbio (<https://sites.google.com/site/geosymbio/>) and in the NSF-funded Dryad digital & open source archived data repository (<http://datadryad.org/>).

Data Access and Re-Use Policies

All of the data generated in this study will be made publicly available upon publication in a peer reviewed journal or within 1 year of the completion of the project. Prior to publication, the community will be given access to unpublished, finalized data upon request, with the caveat that they not publish the data until they can cite our publication presenting the data. In addition to publication in peer-reviewed journals, with all relevant attempts to provide original datasets in open access publication repositories, the data generated from this project will be made publicly available online wherever possible. To ensure accuracy and data tracking, the project will have a specific data use policy including the following:

- User requests require current and valid contact information that will be used by the PI for tracking and documenting data usage.
- Users are required to cite the project publications and acknowledge the NSF as the original funding source.
- Users have the final responsibility for any errors in their external and secondary analyses, while the PIs and project participants will conduct quality control on the primary data and ensure accuracy of the primary data to the best of their abilities.
- The PIs and project participants will not release any private or confidential information to the public, and in-house databases will be password protected.

- The PIs and project participants will retain intellectual property rights, except where released is explicitly required for publication and documentation.

Data Archiving

In the short term, as the data are generated they will be collated and stored on shared internal databases and will be accessible by all involved in the project. Data will be archived in the original data format and also in more common, non-proprietary formats (e.g., tiff, csv, doc) to facilitate future data usage. Data will be mirrored on lab databases at each participating institution. In the long term, genetic data products will be permanently raw-archived with appropriate metadata in the NCBI Sequence Read Archive (SRA). The monthly-resolved coral geochemistry records from the central tropical Pacific, which will be archived at NCDC upon publication. Wherever possible, metadata (data quality control screening, computer code, etc) will be archived with the relevant published dataset. If this is not feasible due to web infrastructure constraints for the major data archive sites, we will archive it on Georgia Tech servers with a web address that is included in the relevant publication. We are firmly committed to open access of our data and methods, and will continue to work diligently to ensure unrestricted access.

We will register all data from this project with the BCO-DMO, either by submitting duplicate data to them, or when possible, helping them to link to datasets that are archived in disciplinary-specific archives such as NCBI-SRA (genomics data) or NCDC (coral and isotopic data). Where BCO-DMO accepts metadata from this project that cannot be archived in specified discipline-specific data archives, we will log such metadata with the BCO-DMO in full compliance with the OCE data sharing policies. In the case of all associated physical data from the upcoming year-long RAPID project, we will upload all data within weeks of its collection, ensuring timely access to data from this coming El Niño event.