

Data Management Plan - Krause

Introduction: Data management will be coordinated by the Principal Investigator. The plan encompasses three areas: use policies, data preservation and archival, and standards. The PI will leverage existing systems wherever possible. All data will be communicated in a timely fashion following the NSF policy.

This project will generate an interdisciplinary dataset including chemical (e.g. radiotracer, benthic flux, O₂ profiles), geological (e.g. authigenic clays in bSi) and biological (e.g. PDMPO and diatoms). Measurements of physical conditions when taking field samples will be done using the available CTD instrumentation aboard the R/V Pelican (see ship time request form).

Data access and sharing policies: Data collected under the project will be made available to the public with as few restrictions as possible. Under these policies, the PI's group plans for publication of most data with metadata, with manuscripts submitted during year 3 of the study and all major results published at a maximum of two years after the completion of the study.

For each sample collected we will generate a suite of data points. Each sample and corresponding subsamples will be assigned a unique identifier which will be associated with sample metadata. This identifier will be used to ensure that data generated in individual labs is associated with the correct sample. All data for each sample will be entered into shared documents (GoogleDocs or DropBox) that will be accessible to all project personnel. This approach has the added benefit of maintaining a copy of the data in the cloud as a backup. Data will be rigorously checked for quality as it is collected, with raw data maintained as well as processed and analyzed data.

Plans for archiving and preserving data: As is required, metadata and data will be contributed to one or more existing catalogs. The PI will work with the Biological and Chemical Oceanography Data Management Office (<http://www.bco-dmo.org>) to archive the data. The PI will work with DISL's Data Management Center, especially the in-house data specialists, Ms. Lei Lu and Ms. Mimi Tzeng, to ensure all data generated and archived conform to the standards of the particular catalog.

The data collected during this research will be archived and made available through public databases following the policies of the Dauphin Island Sea Lab Data Management Center (DMC, dim.disl.org/management_main.cfm). The purpose of this policy is to facilitate and ensure a consistent set of protocols and services for quality assurance, archiving (storage and retention), documentation (such as metadata creation), accessibility, and distribution of data collected by DISL personnel. Metadata will be submitted to DMC to produce FGDC-compliant records. The metadata will be available through the DISL Metadata archive (dim.disl.org/datasets.cfm) and will also be submitted to the National Coastal Data Development Center's (NCDDC), FGDC and Geospatial-One-Stop clearinghouses. The metadata records will provide contact information for people interested in obtaining associated data. Full datasets will be linked to metadata records once the data has been published in peer-reviewed journals. The DMC will coordinate with Information Technology regarding in-house data archives and security measures by providing a written list of directories for nightly backup as well as notification when new databases are added. The DISL server is maintained off-campus, at the University of South Alabama.

Standards and formats to be used for metadata and data: The PI will conform to the metadata standards of BCO-BMO. As much as possible, data will be archived in ASCII format, which is the most flexible and readable over the long term. The PI will archive data in tabular formats that have been proven successful when sharing data.