

Data Management Plan

Overview

The proposed research will generate datasets, models, and model output that will have long-term value to the scientific community, and the team will generate, manage, and archive data using Findable, Accessible, Interoperable, and Reusable (FAIR) qualities to increase the pace of scientific discovery throughout the project and after completion. The proposed work will generate geochemical, process rate, and modeling data. Raw and processed or derived data will be made publicly available via community recognized public databases and peer-reviewed publications. Unless otherwise noted below, all data products, including model outputs, will be prepared and submitted to publicly accessible data repositories in non-proprietary file formats (e.g., csv, txt, png, .tiff) with metadata in machine readable (i.e. xml) ISO 19115-2 standard format for geospatial data. Data repository requirements for additional metadata formats will be met, as outlined below.

We plan on open-access, peer-reviewed publications to be the primary print and digital mechanisms for data and information sharing but will also be presented at scientific conferences, and highlighted in public presentations and media work that is ongoing by the PIs. Validation of results will be enabled by the timely publication of data and metadata in public repositories prior to manuscript publication, or no later than two years after data generation and QA/QC, whichever is first. This follows the recommendations set forth by the Division of Ocean Sciences. When appropriate, and not already included in the main manuscript, data and results will be published as online supplements to the peer-reviewed publications. Those datasets will be assigned static Digital Object Identifiers (DOIs) through respective journals, data repositories or DataCite and linked to NSF award numbers, as permitted. Data management will be coordinated by the PI/PD but involve all co-PIs. Specific archiving plans for data are outlined below.

Data Compliance

We will share and archive data collected as part of this research project in compliance with NSF policy ([NSF11001](#)) and the Division of Ocean Science Sample and Data Policy ([NSF11060](#)) by ensuring (i) all data will be submitted to BCO-DMO within two years of collection; (ii) data, results, and information are made fully publicly available with minimal time delay; (iii) use of existing data and metadata standards and management systems as much as practical; (iv) data interoperability to greatest extent possible; and (v) employment of best practices for data management and policy.

Data generation, storage, and formats

Field/sampling notes and data collected in the field and laboratory on paper data sheets (or Rite-in-the-Rain notebooks) will be scanned or photocopied on the day of collection and entered into tabular spreadsheets. Textual/graphical documents will be made available in at least one of the following formats: plain text, PDF, Microsoft Word, or Microsoft Excel, owing to their widespread use in academia and business. Laboratory analytical and experimental data will initially be recorded on the instruments used for those analyses or on associated laboratory computers as it is collected. Calibration verifications will be stored and archived in the same manner as the data. Field and laboratory data will be stored to desktop computer hard drives and in-house and/or cloud-based servers.

The work plan for the proposed efforts includes the collection of field sediment samples, process rate measurements, field and laboratory notes, and existing data from a range of sources. The project will generate significant quantities of geochemical, rate, and modeling data. None of the data collected or generated during this project will present novel challenges for purposes of archiving and distribution. All field and laboratory/experimental data and accompanying metadata will be stored using the ISO 19115-2 standard at LUMCON and LSU with all final copies stored at LUMCON. Hardware will be based at LUMCON, which has several virtual servers on site and in the cloud, including a network attached server

with ~56 TB set up in a RAID configuration. Existing LUMCON protocols for data backups and disaster management will be applied for all generated data. All analysis scripts (*R*, Python) will be annotated with meta-data describing the original data sources, the purposes of the analysis, the processes used to create the derived data tables and statistical output, who performed the work, and when the final version of the data or analysis script was completed. Model code for reactive transport simulations, including documentation, will be stored, version controlled and distributed through bitbucket, with additional snapshot backups on UGA's cloud storage.

This project will not involve the acquisition of either animal or human subject data.

Data curation

Quality control will be performed on the data produced.

We do not anticipate that there will be any intellectual property issues involved with the data. In the event that discoveries or inventions are made in direct connection with this data, access to the data will be granted upon request once appropriate invention disclosures and/or provisional patent filings are made. The data acquired and preserved in the context of this proposal will be further governed by the policies pertaining to intellectual property, record retention, and data management from the Louisiana Universities Marine Consortium, Louisiana State University, and the University of Georgia.

Backup and archiving

These data and meta-data will be archived at LUMCON, LSU and UGA during the cruises and subsequent experiments, and uploaded to BCO-DMO at the conclusion of the work plan. Metadata and project descriptions will be uploaded annually to summarize ongoing measurements / experiments. The office computers and hard drives mentioned above will be connected to a secure, daily, automatic backup program or service (stored in servers at LUMCON, LSU, and UGA and in the cloud).

Data sharing

We plan to share results of the work principally through open-access peer-reviewed publications. When appropriate, and not already included in the main manuscript, data and results will be published as online supplements to the peer-reviewed publications. We will also share project results with the public via the multiple methods of achieving broader impacts described in our proposal. All publications will include annotated analysis scripts detailing the statistical procedures used in the analyses. We have shared data through publications, project web sites, and outreach activities in the past and are fully prepared to do so in a timely manner for the proposed project.

All geochemical, microbial rate, and modeling data will be submitted in a timely manner to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) at Woods Hole, MA. The PIs will work closely with BCO-DMO personnel to ensure all data generated and archived conform to the standards of the particular catalog. This also includes the reporting of all appropriate meta-data in a timely manner. Upon completion of each cruise, the original underway data will be contributed by the vessel operator to the UNOLS central data repository managed by the Rolling Deck to Repository (R2R) project. Also, R2R will ensure that these data are archived permanently at NODC and/or NGDC as appropriate. Upon any potential publication, data, metadata, and R-scripts associated with individual publications that are not already submitted to a repository will be made available through Dryad (<http://datadryad.org/>). Dryad is an international repository of data underlying peer-reviewed articles in the basic and applied biosciences enabling scientists to validate published findings, explore new analysis methodologies, repurpose data for research questions unanticipated by the original authors, and perform synthetic studies. Dryad is governed by a [consortium of journals](#) that collaboratively promote data archiving and ensure the sustainability of the repository.