Data Management Plan – John D. Kessler, Thomas Weber, DongJoo Joung
University of Rochester

I. Types of Data.

Data management will be led by John D. Kessler, with assistance from DongJoo Joung and Thomas Weber. Products of the proposed project will fall into four categories: (1) oceanographic data collected by in-situ or underway oceanographic devices, (2) information related to the concentration and isotope analyses conducted on samples returned to the laboratory, (3) codes for training and application of Artificial Neural Network (ANN) models, (4) gridded datasets of methane supersaturation and flux produced by the model. The PI proposes the following Data Management Plan:

II. Data access and sharing.

For the oceanographic data, we will follow procedures set by the Biological and Chemical Oceanography Data Management Office (BCO-DMO), which has a sophisticated data management system and personnel experienced with standards and data curation. For our work in the UNOLS vessels, the underway data will be automatically submitted to Rolling Deck to Repository (R2R) database (http://www.rvdata.us/). The PI will also maintain a copy of all underway data (Meteorological, CTD-bases, navigational, ADCP, bathymetric and underway sensors). Data unique to this study will be linked to the archived source of the appropriate meta-data at the time of publication, in addition to presenting pertinent metadata in publication.

For the data collected in direct support of concentration and isotopic analyses, we will maintain the data in laboratory notebooks and in electronic form, with regular automated backups to both backup drives and cloud sources. This data will include concentrations and isotopic compositions of methane, and concentrations of nutrients, dissolved oxygen, and chlorophyll, as well as ancillary metadata. This data will be made available to the scientific community through publication and the Biological and Chemical Oceanography Data Management Office (BCO-DMO) database (https://www.bco-dmo.org/). All methodological details of the sample collection and analyses will be included in the publication and/or database records. Anticipated metadata includes location, time, depth, temperature, oxygen concentration, salinity, and density, and possibly other data.

Original codes developed for all stages of the ANN model workflow (including data preparation, model training, and model application to produce gridded datasets) will be housed on the University of Rochester's off-campus Bluehive cluster. A generalized version of these codes will be prepared and submitted to an online code-sharing archive (e.g. GitHub). This will be designed for accessibility to other researchers, allowing them to apply our methods to their own datasets, with minimal modification. The large gridded datasets of methane supersaturation and ocean-atmosphere flux produced by our ANN model will likely exceed the size limitations of most online archives. They will be housed on storage servers maintained by the University of Rochester's Center for Integrated Research Computing (CIRC), where Weber owns 40TB of dedicated storage space. These datasets will be made available for transfer (by secure FTP protocol) upon request.

Research results will be disseminated to the broader research community also through journal publications and presentations at conferences. We have and will continue to publish data as supplemental online material, as this provides a permanent archive with world-wide access. Published papers will be made available in PDF format, per the websites of the specific professional journals. There will be no restrictions on access to these data, other than the journals' copyright restrictions.

III. Security.

The original files for digital cruise and laboratory data will also be maintained on dedicated computers in the PIs' laboratories and offices. These files are backed up daily to a remote hard-drive and cloud service. The office is locked when the PI is away and adequate security measures are taken to maintain the integrity of the data, both with respect to the electronic repositories and personal computers, per standard Information Technology best practices. All computers are password protected and run over encrypted servers at the University of Rochester.

Model codes and output will be stored on the Bluehive and CIRC storage servers, which are protected by encrypted firewalls and maintained by a dedicated staff that includes digital security experts.

IV. Responsibility.

PI John D. Kessler will be responsible for data management throughout the project and subsequently, with assistance from DongJoo Joung and Thomas Weber.