

## Data Management Plan

**1. Types of Data Collected:** Four types of data were collected: 1) molecular sequence data, 2) quantitative PCR (qPCR) and cell counts 3) physiological rate data such as cell specific and bulk uptake rates for N<sub>2</sub> and Carbon fixation, 4) electron micrographs of diatoms, and nanoSIMS images of element (stable) in individual symbiotic diatoms. Data collected will be used to generate primary publications in the peer reviewed literature including relevant primary data presented in supplemental tables.

**2. Standards to be Used:** All analytical determinations for environmental data as well as for rate measurements which are derivative from analytical data are calibrated against known standards. For qPCR assays, standards are made from linearized plasmids of the target genes, i.e. *nifH*. For stable isotope measurements, the samples analyzed are calibrated against NIST Standard Reference Materials (IAEA-N1, IAEA-N2, IAEA-N3, USGS-40, and USGS-41).

**3. Policy for Access:** After primary publication, we will freely share our data with any legitimate user. As mentioned, we routinely publish our primary data as supplemental tables for general access.

**4. Policies and Provisions for Re-use:** As per 3.

**5. Plans for Archiving Data:** Data sets will be submitted to national data centers Biological and Chemical Oceanographic Data Management Office (BCO-DMO), all data set related to nitrogen fixation will be added to the current database for N<sub>2</sub> fixation stored in PANGAEA (<http://doi.pangaea.de/10.1594/PANGAEA.774851>), and sequence data will be submitted to the national center for biotechnology information ([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)).