Data Management Plan

(Barbeau and Dupont)

Collaborative Research: Multiple Analytical Window Electrochemical Techniques and Meta-Omics Applied to Studies of Iron Recycling and Iron-Binding Ligands in the Ocean

1 **Data Policy Compliance**

The project will comply with the data management and dissemination policies described in the *NSF Award and Administration Guide* (AAG, Chapter VI.D.4) and the *NSF Division of Ocean Sciences Sample and Data Policy*.

2 **Fieldwork Planning**

The PI will plan sampling from the SIO Pier and local small boat sampling in consultation with co-PI Dupont and the graduate students involved in the project. Actual pier and small boat sampling events will be recorded on paper logs (scanned into PDF documents).

3 **Description of Data Types**

The project will produce several observational and experimental datasets, described in the list below. Observational and experimental data will be collected from the SIO Pier and from small boats deployed from the Pier. Data may also be obtained from cruises of opportunity, as described below.

Observational Datasets:

SIO Pier samples: Time series samples will be collected from the SIO Pier for iron speciation (dissolved iron concentration and concentration and binding strength of organic iron-binding ligands); extractable chl, macronutrients, POC/PON, and microscopic counts. Sampling events (dates, times) will be logged manually. File types: Raw (.con, .hdr, .hex, .bl) and processed (.cnv, .asc, .btl) and ASCII files. Repository: BCO-DMO. DNA/RNA will also be collected and data will be handled through JCVI, see below.

Experimental Datasets:

Incubation studies: Incubations will be carried out with water and net tow samples collected from the SIO Pier and local waters around SIO to study the evolution of the iron-binding ligand pool during organic matter remineralization. Dataset will include iron speciation, POC/PON/TOC, nutrients, extracted chl, and microscopic cell counts. File types: Excel file(s). Repository: BCO-DMO. DNA/RNA will also be collected and data will be handled through JCVI, see below.

Potential Cruises of Opportunity Datasets: During the time course of this project, opportunities may arise for the use of the Sorcerer II as a sampling platform for gathering observational data. This will include total dissolved iron and iron speciation samples (File types: Excel files. Repository: BCO-DMO). Processed (.csv) and raw depth (.con, .hex) profiles from a SeaBird CTD package equipped with fluorometer and oxygen sensor will be provided (Repository: BCO-DMO). Additionally, an opportunity may arise for leveraging a California Current Ecosystem Long Term Ecological Research (CCE LTER) Process cruise in the Southern California Current System to obtain observational and experimental samples in support of this project. All CCE LTER associated observational and experimental data (including total dissolved iron and iron speciation, extractable chl, macronutrients, POC/PON, and microscopic counts, ancillary hydrographic data from CTD casts and ship's underway system, event log) will deposited in the CCE LTER primary data catalog, Datazoo (http://oceaninformatics.ucsd.edu/datazoo/). Any DNA/RNA data obtained for this project on Sorcerer II or a CCE LTER process cruise will be handled via JCVI, see below.

Meta-omics Datasets: As part of the proposed work, ~120 Gbp of shotgun metatranscriptomic data, ~120 Gbp of shotgun metagenomic data, and 22.5 Gbp of loci-based metagenomic data will be generated. The timely deposition of these data and the associated experimental and environmental data in

several public databases has been accounted for in the JCVI budget though costs for data processing, deposition, and storage (IT equipment). Prior to the submission of publications describing the data, the raw Illumina reads will be deposited in the NCBI Sequence Read Archive, while the assembled transcripts will be deposited in the NCBI Transcriptome Shotgun Assembly Database. JCVI is an active participant in the Genome Standards Consortium, which issues policy recommendations on data deposition such as the Minimum Information for Metagenomes Standards (MIMS). In accordance with these policies, experimental details and data-processing information will be included with data depositions, along with the measured environmental parameters. If new scripts for statistical analyses are developed, it will be in the open-source R programming language and they will be distributed free-of-charge. The JCVI budget also includes the cost of long-term (10+ years) data storage arrays, which provides a failsafe should other public repositories lose funding.

4 **Data and Metadata Formats and Standards**
Field observation data will be stored in flat ASCII files, which can be read easily by different software packages. Field data will include date, time, latitude, longitude, cast number, and depth, as appropriate. Metadata will be prepared in accordance with BCO-DMO conventions (i.e. using the BCO-DMO metadata forms) and will include detailed descriptions of collection and analysis procedures. Any CCE LTER data associated with this project will be deposited in Datazoo with appropriate metadata according to LTER network quality standards and best practices.

5 **Data Storage and Access During the Project**
The investigators will store project data (including spreadsheets, ASCII files, images, and PDFs of scanned logs) on laboratory computers that are backed up regularly.

6 **Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution**
Primary data sets produced by project participants will be made available through the BCO-DMO data system within two-years from the date of collection. The project investigators will work with BCO-DMO data managers to make project data available online in compliance with the NSF OCE Sample and Data Policy. Data, samples, and other information collected under this project can be made publically available without restriction once submitted to the public repositories. Data produced by this project may be of interest to chemical and biological oceanographers, and climate scientists interested in the role of biogeochemistry in the global climate system. We will adhere to and promote the standards, policies, and provisions for data and metadata submission, access, re-use, distribution, and ownership as prescribed by the BCO-DMO Terms of Use (http://www.bco-dmo.org/terms-use). Any CCE LTER data associated with this project will be made available to the public according to LTER network quality standards and best practices.

7 **Plans for Archiving**
BCO-DMO will also ensure that project data are submitted to the appropriate national data archive. The PI will work with BCO-DMO to ensure data are archived appropriately and that proper and complete documentation are archived along with the data. Any CCE LTER data associated with this project deposited on Datazoo will be submitted to the LTER network data system for long term archiving. JCVI will archive metagenomic and metatranscriptomic data.

8 **Roles and Responsibilities**
Barbeau will coordinate the overall data management and sharing process and will submit the project data and metadata to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) who will be responsible for forwarding these data and metadata to the appropriate national archive. Barbeau will also work with the CCE LTER data manager as necessary if the CCE LTER cruise of opportunity is an option. co-PI Dupont will be responsible for submitting metagenomic and metatranscriptomic data.