DATA MANAGEMENT PLAN

In accordance with NSF Award and Administration Guide (Chapter, VI.D.4), the following measures and policies will be undertaken as part of the proposed research:

1. Types of data, samples, and other materials to be produced in the course of the project.

This project will produce the following types of data from samples of Lake Erie:

- i. Biological data:
  - a. Sequence of DNA and cDNA
    - i. Raw files.
    - ii. Fasta format:
      1. Raw sequence reads
      2. Consensus sequence of assembled contigs
    - iii. Gene annotations
  - b. Rates of bacterial and primary production
- ii. Physical data: latitude/longitude, temperature, depth, and conductivity.
- iii. Chemical data: concentration of nutrients, hydrogen peroxide, dissolved organic matter, dissolved oxygen

2. Standards to be used for data and metadata format and content.

As part of the proposed research, summary and data spreadsheets will be made compiling the samples collected and analytical results. The project summary spreadsheets (sample types collected, sample numbers, etc.) that list the samples and data that were collected or generated as part of the project will be posted to the PIs' website in the "Data Download" section designed to facilitate dissemination of project data. Data generated by this project will be placed in archival web services. At the present time, there is not a single, non-migrating archive for all types of data that will be generated by the proposed research.

The above data will be stored on local computers and backup servers at the University of Michigan.

Data of type (i.a) will be stored as raw image files, raw sequence reads (fasta text files), consensus sequences (fasta text files), and gene annotations (MS excel spreadsheets spreadsheets). Data of type (i.b), (ii), and (iii) will be stored in as raw image files or as MS excel spreadsheets (*.csv format) and in hard copy lab notebooks.

3. Methods and policies for providing access and enabling sharing.

Data storage redundancy will be conducted following best practices and under the guidance of the scientific computing administrator for the University of Michigan Department of Earth and Environmental Sciences.

Until finalized and archived, data will be made available by e-mail request and in the "Data" section of the PIs' websites to facilitate those requests or to post the data directly.

4. Provisions for re-use, re-distribution, and the production of derivatives.

As appropriate or as possible, samples will be made available for re-analysis or for new analyses by other groups. Outreach materials created as a part of this project will be digitally archived for possible future redistribution.

5. <u>Methods for archiving and preserving access to data and materials.</u>

i. **Potential users.** The major users will be the microbial ecology community; primarily microbial ecologists but researchers from other fields are also anticipated. The datasets will also be used for educational purposes in graduate and undergraduate classes at the University of Michigan.

ii. **Timing of release of data.** Data of type (i) will be **deposited to public databases such as NCBI and IMG upon publication or two years after their generation, whichever comes first**. Data of types (ii) and (iii) will be published in peer-reviewed journals, will be made available on the project web site and at the Biological and Chemical Oceanography Data Management Office metadata database, and will be further available upon request.

iii. **Data attribution.** Data will be appropriately attributed through authorship of publications and public database entries; datasets will be assigned *doi* numbers for archival attribution.

iv. **Dissemination of results.** Significant research results will be written up and submitted for publication rapidly. Whenever possible, all of the data used in the publications will be published as part of the paper itself or via a journal's Data Repository. The data will also be made available by e-mail request and the "Data" section of the PI's website to facilitate those requests or to post the data directly.