Data management Plan

1. Describe the data that will be collected, and the data and metadata formats and standards used.

Mass spectrometry data from metabolomics will be archived in gzipped mzML format. Although no formal standards exist for metadata requirements for proteomic and metabolomic data, MIGS/MIMS/MIENS standards for metadata will be captured where possible.

Cell cultures isolated during the project will be stored in a liquid nitrogen Dewar as part of the OSU High Throughput Culture Collection and will be designated with a unique identifier. Isolates will be maintained in two different Dewars for redundancy and will be made available through the OSU Office for Commercialization and Corporate Development upon completion of a Uniform Biological Materials Transfer Agreement as detailed on the Giovannoni Laboratory webpage (HYPERLINK "http://giovannonilab.science.oregonstate.edu/htcl-contributions-science-and-industry" http://giovannonilab.science.oregonstate.edu/htcl-contributions-science-and-industry).

Genome sequence data (genomes from cultures, single-cell genome data and metagenomic data) will be collected and curated according the Genomic Standards Consortium (http://gensc.org/gc_wiki/index.php/Main_Page) Minimum Information about a (Meta)Genome Sequence (MIGS/MIMS/MIENS) standards (http://gensc.org/gc_wiki/index.php/MIGS/MIMS). These outline a standardized format for the minimum information required to accurately describe genomic and metagenomic data, including metatdata, with the goal of facilitating inter-study comparisons and transparency (Field et al. 2008).

For maximum reproducibility and transparency of bioinformatic analyses, all processes from raw-data to publication results will be documented as runnable manuscripts, either in IPython Notebook or as R Sweave projects. These will be made publicly available through the Giovannoni Lab GitHub repository (HYPERLINK "https://github.com/organizations/giovannoni-lab" https://github.com/organizations/giovannoni-lab) and through the Giovannoni Laboratory webpage (HYPERLINK "http://giovannonilab.science.oregonstate.edu/data-analysis-programs" http://giovannonilab.science.oregonstate.edu/data-analysis-programs) upon publication of results.

2. Describe what physical and/or cyber resources and facilities (including third party resources) will be used to store and preserve the data after the grant ends.

All standard oceanographic data will be hosted by the Biological and Chemical Oceanography Data Management Office (BCO-DMO). Metadata of cruise activities will be generated on a quarterly basis and datasets will be transmitted to the BCO-DMO within one year of their collection.

Raw sequence data will be deposited in Sequence Read Archive (SRA) in FASTQ format, with relevant contigs from assemblies deposited in the Genbank as Whole Genome Shotgun (WGS) projects or complete genomes as appropriate. Publication of datasets will be accompanied by either a Data note in GigaScience (HYPERLINK "http://www.gigasciencejournal.com" http://www.gigasciencejournal.com) describing the dataset, or as genome announcements in Standards in Genomic Sciences (HYPERLINK "http://www.standardsingenomics.org" http://www.standardsingenomics.org) where appropriate.

MS/MS data from proteomics and metaproteomics analyses will be made publically available by deposition in the Proteomics Identifications Database (PRIDE, HYPERLINK "http://www.ebi.ac.uk/pride" http://www.ebi.ac.uk/pride).

3. Describe what media and dissemination methods will be used to make the data and metadata available to others after the grant ends.

Links to all data storage are maintained centrally on the Giovannoni Laboratory webpage: HYPERLINK "http://giovannonilab.science.oregonstate.edu/" http://giovannonilab.science.oregonstate.edu/. The Giovannoni Laboratory webpage is maintained by the OSU Central Web Services office, which ensures that data remain accessible and available after grant funding has ceased.

Metaproteomic and metagenomic datasets will be deposited in the appropriate public repository described above and will be accompanied by a Data Note in GigaScience (HYPERLINK "http://www.gigasciencejournal.com" http://www.gigasciencejournal.com), to promote maximal re-use and provide easy access to accompanying rich metadata.

4. Describe the policies for data sharing and public access (including provisions for protection of privacy, confidentiality, security, intellectual property rights, and other rights as appropriate).

All data are made available immediately upon publication of results, and with rare exceptions all sequence data are released by standard embargo dates, as mandated by agencies. Unpublished data of other types are released at the discretion of the P.I., with the intent of maximizing the dissemination of data via peer-reviewed formats.

The Giovannoni Laboratory provides live bacterial isolates and to academic research groups around the world (HYPERLINK "http://giovannonilab.science.oregonstate.edu/ htcl-contributions-science-and-industry" http://giovannonilab.science.oregonstate.edu/ htcl-contributions-science-and-industry), usually at no cost to the requesting institution. The Giovannoni Laboratory fully supports and is committed to the timely distribution of scientific materials and data for the advancement of science.

5. Describe the roles and responsibilities of all parties with respect to the management of the data (including contingency plans for the departure of key personnel from the project) after the grants ends.

Project personnel are responsible for maintaining notebooks and computer files. The PIs are responsible for storage of notebooks and computer files after the departure of personnel. As outlined above, most data are stored on independent, long-term, third-party server hosts that make them continuously available.