

Data management plan

NSF Antarctic Research: Viral control of microbial communities in Antarctic lakes: from satellite to giant viruses

I. Data Policy Compliance

Data generated as part of this project will comply with the NSF Division of Ocean Sciences Data and Sample Policy dated May 2011 and online at: www.nsf.gov/pubs/2011/nsf11060/nsf11060.pdf. In addition, we will adhere to the Bigelow Laboratory for Ocean Sciences institutional data management plan. Policies within the plan will continue to develop as centralized computational and/or storage resources are developed. Genomic data: We will comply with Minimum Information about a Genome Sequence (MIGS) specification as defined by the Genomic Standards Consortium (GSC). All data quality will be assured through analysis of replicate samples and proper accounting of standards and controls.

II. Types of data produced during this project

a) Environmental metadata associated with individual field samples: i) Quantitative flow cytometry of autotrophic, heterotrophic and virus communities in the water; ii) Physicochemical parameters when collected (e.g. seawater temperature, salinity, nutrient concentrations).

b) Molecular and sequence data: i) Virus genomes, metagenomes, digital PCR.

III. Mechanisms for access and sharing of data

a) Lab books: initial documentation

b) Digitally: i) Bi-monthly scanning of initial documentation in lab books to provide digital records; ii) Back up on local server.

c) Databases (for wider distribution): i) Virus, bacteria, and phytoplankton FCM enumeration data and related environmental sample data will be provided as part of scientific publication and/or upon request to project PIs; ii) GenBank public archive, for nucleic acid sequence information in accordance with the NSF Sample and Data Policy;

d) Outreach activities through the Bigelow Laboratory Communications office.

e) Scientific paper publications, presentations at national and international workshops and conferences.

f) Annual and final reports on data and research product to NSF

IV. Plans for archiving

Long-term preservation shall be facilitated using Bigelow Laboratory's robust, multi-tiered data backup and retention strategy. Snapshots are performed on an hourly basis (for 72 hours), daily basis (for 30 days), and weekly basis (for 52 weeks). Storage volumes are also synchronized nightly to a secondary storage tier, should the primary tier experience a catastrophic failure. Lastly, volumes that are deemed to be ready for permanent archive are encrypted and stored on Amazon Glacier. Data submitted to NODC are maintained in perpetuity.