

Data Management Plan

EAGER: Persistent virus infections in marine phytoplankton.

• 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. Such policies will be for the purpose of safeguarding data assets with the objective of minimally reducing Laboratory liability with regard to agreed upon outcomes from publically grant-funded scientific projects.

• Types of data produced in the course of the project:

a) Environmental metadata associated with individual field samples:

i) Quantitative flow cytometry of autotrophic, heterotrophic and virus communities.

ii) Physicochemical parameters when collected.

b) Phytoplankton and virus isolates for live curation:

i) Original sample metadata.

ii) Tracking isolation optimization.

iii) Characterization data:

(1) Growth and propagation.

(2) Host range data (for viruses).

(3) Electron microscopy.

c) Molecular, genomic and sequence data

i) Genome sequencing.

ii) Ribosomal genes.

iii) Virus diagnostic marker genes.

iv) Other taxonomic genes.

v) qPCR data.

d) Broader Impacts:

i) Outreach efforts

ii) Student tracking

iii) Economic impact statements

• Standards to be used for data and metadata

a) Genomic data:

We will comply with Minimum Information about a Genome Sequence (MIGS) specification as defined by the Genomic Standards Consortium (GSC)

(http://gensc.org/gc_wiki/index.php/Main_Page).

• Mechanisms for access and sharing

a) Lab books

i) initial documentation

ii) Periodic scanning to provide permanent digital record

b) Digital

i) Initial documentation (laptops)

ii) Digital gel images

iii) Micrographs

iv) Sharing locally using Google docs or Dropbox

v) Backed up on local server and external hard drives.

c) Databases (for wider distribution)

i) BCO-DMO

ii) CAMERA

iii) NCBI

iv) Single Cell Genomics Center (SCGC) Laboratory Information Management System (LIMS)

v) National Center for Marine Algae and Microbiota (NCMA) MySQL database.

d) Lab project page web site (advertise project information and availability of samples)

e) Biological specimen distribution (NCMA: <https://ncma.bigelow.org/>)

f) Outreach activities through the Bigelow Laboratory Communications office.

g) Publications, workshops and conferences.

h) Bigelow Laboratory IP Policy: In accordance with the 1980 Bayh–Dole Act, all data generated as a result of research performed at Laboratory, or with Laboratory funds, shall be the intellectual property of the Laboratory and the scientists overseeing the research. As such, the Laboratory will retain an implied copyright for these data.

i) Reporting:

i) Progress on data and research product sharing will be addressed specifically in the annual report to NSF.

ii) A complete account of data policy compliance will be recorded in the NSF Final Report.

• **Policies and provisions for re-use**

a) Embargo periods during which data may be inaccessible by the public will be the discretion of the overseeing SRS and the Executive Director of Bigelow Laboratory. These will most often correspond with publication dates, etc. The right-of-first-use for original data will be retained by the overseeing PI, except as the result of extreme circumstances as judged by the Executive Director and the larger Bigelow Laboratory Senior Research Scientist Committee.

b) All data will be submitted to BCO-DMO (or other nominated databases) no later than 6-months after collection of field- and lab-work.

c) Synthesis products will be made publically available no later than 6-months after the project funding expires or as soon as data is published.

• **Plans for archiving**

a) **Long-term strategy.** One year after a project has been completed; arrangements shall be made to transfer data from short-term storage to a long-term archival system. The data from different projects will be maintained in separate file structures with the SRS's hierarchical file system. Long-term preservation shall be facilitated using a strategy that has been endorsed by the Laboratory IT specialist.

b) **Length of archival.** Data will be kept in long-term storage for at least five years, or until it has been successfully uploaded to and made publically available by a nationally or internationally funded database specific to that data.