https://www.bco-dmo.org/project/774168
Data Management Plan

We have read and agree to abide by NSF guidelines and award conditions for scientific conduct and data management. Results, data, and collections will be made available to qualified researchers upon request, provided that the quantities and time requirements for compliance do not compromise our research objectives. We will also address data sharing issues in annual and final reports.

Lead Investigator: Matthew B. Sullivan
Institution: Ohio State University
Collaborators: Adam Rivers (USDA), Steven Hallam (U British Columbia, Canada), Alex Culley (U Lavalle, Canada), Tara Oceans Consortium collaborators
Project: Ecology and biogeochemical impacts of DNA and RNA viruses throughout the global oceans
Solicitation Info: NSF PD 98-1650
Submission Date: Feb. 15, 2018

Project overview: A global oceanographic research expedition has been completed and data generation, curation and management has been centrally funded by the *Tara* Oceans Consortium. This proposal seeks funding to support sample preparation for new data generation (49 dsDNA viromes, 139 RNA viromes, bacterial and viral abundance data), analyses of viral and relevant sequence datasets to establish RNA virus reference genomes and evaluate viral 'activity', and analyses of viral and contextual data to look at ecological drivers of community structure.

General data management and sharing overview: An NSF BCO-DMO page will be created to describe this project, but it would then point to the Consortium's use of the PANGEA database for access to all data and contextual resources from the expedition. Since this project is largely built from data already collected and generated through the Tara Oceans oceanographic expedition, most of the data needed for the project is already available as raw data through the PANGEA database for meteorological, oceanographic, biochemical and plankton morphological data, and through the ENA repository for molecular data (http://www.ebi.ac.uk/ena/submit/tara-oceans-checklist). In additions to making our data available to the Consortium through these avenues, we intend to make all data that is annotated and curated through this project publicly available through 'iVirus'. The iVirus project represents a set of dozens of software tools (apps) and related public datasets that are dedicated to viral ecology and built upon the NSF-funded CyVerse Cyberinfrastructure. iVirus was collaboratively developed by Sullivan and Assistant Professor Bonnie Hurwitz (U Arizona), and Sullivan continues development and resource deposition there. The few data generated by this project will be handled in the same manner via BCO-DMO update, PANGEA raw data release, and curated data products and documentation through iVirus.

Locally, we now routinely work with data at this scale and have automated back-up procedures at multiple levels (private compute cluster, university and Ohio State high-performance compute cluster). All curated data and tools / apps will be made available to the Consortium immediately and to the research community in conjunction with the peer-reviewed publication that describe the materials if not before.

Data archives: Viral metagenome raw data will be stored in NSF-funded CyVerse facilities through the iVirus portal in addition to the ENI. Curated viral genomes and their associated annotation (e.g. predicted host, detection in microbial metagenomes, etc) will be made available as a separate database at iVirus, in both fasta and genbank file formats so that a user can easily use these as new references. Other curated data products (e.g. co-occurrence networks, taxonomic networks, etc) will also be made available through iVirus. All records will be cross-referenced to sample data, and will follow the naming guideline of the *Tara* Oceans expedition.