

## DATA MANAGEMENT PLAN

### **We have read and agree to abide by the Division of Ocean Sciences Data and Sample Policy.**

Metadata associated with this proposed research, including information on sites, experiments, and data collected (e.g., date, time, location, experimental treatments and maintenance, and environmental variables measured) will be archived and made accessible to researchers and the public on the Northeastern University Marine Science Center website. Furthermore, we will provide archives of all environmental data collected at our three study sites (e.g., water and air temperature) for researcher and public download on our website. Data summarizing our experimental results, which may be useful for syntheses, meta-analyses, or projects evaluating similar research questions, will be shared if requested.

Our project will also produce various types of genetic data, which we are committed to making open access. All of these genetic data will be archived in the NCBI databases once they are fully annotated. We will use our lab website as a clearing-house for the data. Our website will include updates on recent data additions and links to our data on the website and in traditional data repositories (e.g., NCBI). The plan for each data type is briefly outlined below:

- (1) **Illumina V6\_rRNA sequencing data:** All of the 16s rRNA sequencing data will be archived and backed up on servers at the Marine Science Center. All of the 16s rRNA sequencing reads will be submitted to NCBI RSA archive and the assembled and annotated 16s OTU sequences will be posted on the NCBI databases. We will make the data available on our website (even in the early stages of the data processing) by posting early annotations for download as well as the raw 16s rRNA sequence data.
- (2) **Genomics:** Genetic samples (Preserved tissues and DNA extractions) will be deposited into -80°C and -20°C chest freezers at the Marine Science Center, Nahant, MA and will be made available to researchers upon request.
- (3) **Other:** All other metadata (e.g., temperature, location, pH, disease state) will be made available via a project link on our website. The link will be updated frequently to highlight how the data deliverables are proceeding and where they can be obtained. In addition, all environmental data will be deposited with the Biological and Chemical Oceanography Data Management Office (BMO-DMO).

All computer code used to create and analyze the mathematical model (Aim 2) will be distributed to the widest possible audience via a new open-source R package. This package will be posted on GitHub (<http://github.com/tgouhier>) and the Comprehensive R Archive Network (CRAN: <http://cran.us.r-project.org/>). This package will also include the spatially-explicit and temporally-resolved field data on coral disease incidence and associated microbial community structure as a fully annotated, flat file (see Gouhier's R packages for examples of this; **biwavelet**: <http://tgouhier.github.io/biwavelet/>, **synchrony**: <http://tgouhier.github.io/synchrony/>). The availability of such spatially-explicit data will promote our understanding of coral disease dynamics by allowing us to discriminate between alternative drivers of outbreaks using the next generation of epidemiological models.