DATA MANAGEMENT PLAN   (Responsible laboratories: Segrè, Sher)

1. Products of Research

The project will collect and analyze the following data:

- **Sample records (source/date/host etc.):** created by PhD students or postdoc (Sher) in lab notebook; Notebook is written with data-storing pen (Livescribe) and backed up daily onto lab computer as a searchable pdf. Lab computer is backed-up daily to project folder on the off-site, server-based departmental storage "N:").

- **Experimental measurements:** These include flow cytometry, FRRF, UPLC chromatograms, metabolite concentrations, oxygen electrodes etc. They also include CTD data from cruises; stored indefinitely on hard drive of relevant measuring system, and duplicated to N: Prints will be placed in the project notebook (Sher) and CTD data will be made openly available immediately after collection on the THEMO website.

- **Transcriptome and proteome data:** Ilumina sequences created by the NGS centers used for sequencing (we typically use the Technion, Weizmann Institute of Science or University of Chicago); raw data downloaded to lab server and backed up on additional external hard drive. Final analyzed data transferred to N: (Sher)

- **Protocols & novel techniques:** hardcopies in project-specific binders (Sher), electronic copies on N: techniques recorded in lab notebooks; protocol modifications annotated directly on hard- and electronic copies (Sher).

- **Theory and models:** theoretical developments in peer reviewed publications; dissemination of algorithms and numerical model codes ensures reproducibility; open-source codes to be posted at mitgcm.org (F).

The project will create the following materials:

- **Culture samples** (Sher): collected (filters or Eppendorf test tubes), maintained in appropriate buffer at -80^0C.

- **Extracted DNA, RNA, proteins, macromolecules** (Sher): in 2mL tubes will be stored frozen.

- **Illumina Library preparations** (Sher): stored in 0.5mL tubes, frozen for duration of project.

- **Samples for nutrient analyses** (Sher): in 50mL tubes, frozen until analysis

  .

2. Data Storage and Preservation

Dry storage: Hardcopy data, lab notebooks and protocols will be archived in a project-specific dry storage boxes (Sher).

Freezer storage: Culture samples stored as described above. Locations of each sample will be noted in dedicated freezer charts (hardcopy sheets in binders). At the completion of the project, where practical, key samples (e.g. Libraries, RNA samples) will be maintained indefinitely in deep-freeze. Expired material (e.g. macromolecular samples) will be discarded. Duplicate samples from field campaign (THEMO) maintained in two separate -80 freezers as community archive (Sher).

Digital data: Daily 'working copy' of data on local machines, with daily (minimum weekly) duplication of those data to a project-specific data folder on an off-site, RAID-protected, backed-up online storage service (N: drive). At project end, all models will be made available as open-

source, annotated code through the Sher and Follows lab websites and/or Github (Segrè). Hardcopies will be printed out and held in dry storage. Publicly archived data will be stored according to individual database policies.

- **Transcriptomics data**: archived and made publicly available at NCBI's Sequence Read Archive (SRA) and the Sher lab website, and connected when relevant to the Chisholm lab ProPortal website.

- **Protocols & novel techniques:** published as Methods and shared freely on request.

3. Data Formats and Metadata

- **Culture samples:** in laboratory notebooks and in project sample database (Excel), including all relevant experimental data (nutrient concentrations, FRRF results, cell numbers etc).

- **Flow cytometry data:** standard FCM formats. File names will contain sample ID that will cross-reference to lab notebook.

- **Transcriptome sequence data**: sequences in standard FASTQ text file format, FPKM data as Excel spreadsheets.

4. Data Dissemination & Policies for Data Sharing and Public Access

The two PIs strongly believe in a policy of network building and collaboration through the sharing of data and materials. Each have publicly accessible websites where they provide periodic, plain-English updates on their research; the websites will also provide a gateway for freely sharing data. The Model code (implemented in MATLAB, Python and/or FORTRAN) will be made available as open-source code through these websites or Github (Segrè).

5. Roles and Responsibilities

Specific people or labs who will create, handle and store the data and samples are given in the sections above. The grant budget provides funds for local hard drives to store and transfer data in the short term. Daily backup of lab computers to N is handled automatically by software, and supervised by the Charney School of Marine Sciences IT office.