Data Management Plan

Overview:
This project will generate significant quantities of data from several laboratories and field locations. This data will consist of 1) terrestrial data from supraglacial meltwaters and proglacial meltwater rivers, and 2) marine data from fjords. We will adhere to the dissemination and sharing of research results as detailed by the NSF Data Management Plan throughout the project duration.

Roles and responsibilities:
The project PI, Dr. Jon Hawkings, will take ultimate responsibility for the data management. It is the responsibility of Hawkings to ensure that all project participants are fully informed and appropriately trained in proper data management and always uphold NSF requirements. All members of the project will take individual responsibility for archiving and backing up the data they collect. Data management responsibilities will be clarified at the start of the project in the first all hands meeting, with time set aside for data management discussion in all other project meetings. The NSF Arctic Data Center will provide archival, access and metadata services for the project.

Types of data produced:
The data generated during this project will be from in situ field collection and lab-based analysis of collected samples.
Physical samples: Samples collected in the field will be labelled with the sample location, type, time and date of collection. These details will also be logged in a field notebook (waterproof paper and ink/pencil), along with sampler name and how the sample is preserved (e.g., "ambient", "frozen", "acidified with HCl to 0.2 % v/v"), with these details input into a computer spreadsheet (.xlsx) as soon as logistically feasible. Samples taken during fjord work will additionally be labelled with a station code (pertaining to the location) and approximate depth (as determined by the idealized sampling depth). Corrected depths will be determined using the CTD attached to the Kevlar line after the sensor data has been downloaded (see below). All spreadsheets will be shared securly amongst the project team (via Penn Box drive). Upon return to labs, the location of samples will be updated and all data double checked against hand written notes.
Observational data: Weather conditions, GPS locations, elevation and other observation data will be recorded in field notebooks and then input into spreadsheets as soon as logistically feasible. Photos of field locations will be acquired, and the image code noted to link with the sampling information. We will be using Garmin GPS with InReach technology during fieldwork, meaning that GPS locations stored in the field are automatically added to web servers via satellite communication, and held there in perpetuity (as an additional back up).
Sensor data: Field sensors will be used to collect high temporal resolution electrochemical datasets. For terrestrial work we will collect water quality information from two sondes – one handheld for spot sampling and one mounted in the river for continuous measurements. This sensor data includes water temperature, water depth, specific conductivity, pH, oxygen and turbidity. For fjord work we will collect information on the water column physical-chemical properties using a CTD profiler. This fjord sensor data will include high resolution (8 Hz) water temperature, salinity (conductivity), depth, chlorophyll, turbidity, photosynthetically active radiation and dissolved oxygen. Data will be download from sensors packages (sonde and CTD) immediately (handheld sonde and CTD) or every day (river mounted sonde). This data will be stored locally (i.e., on field portable laptops) until internet connectivity is available, when they will be uploaded to a share cloud drive (Penn Box account) and filed appropriately (file name to have standardized YEAR-DAY-TIME-LOCATION under a folder named after the sensor package). Periodically throughout the field season, data will be assembled into master spreadsheets.
Analytical data: Large quantities of data will be generated during laboratory analyses (see Table 1 of Project Description for all analytes and methods). This data will include raw instrumental data (e.g., counts, voltage response, absorption), calibrations to know standards, and processed data (calibrated

data). Analytical metadata (e.g., sample name, autosampler position, volume, method etc…) will be recorded in laboratory notebooks and uploaded to instrument run spreadsheets and spreadsheets (.xlsx) at the earliest opportunity. Methodological protocols will be details in word processor documents and shared between team member via the Penn Box drive (as above). Where similar analyses are being done in two different labs, labs will confer beforehand to standardize methodology, and/or indictors of success. Raw instrument data will be named YEAR-DAY-TIME-LOCATION-ANALYTE-OPERATOR-"RAW". Processed data will be named similarly, but with "PROCESSED"-[version number] instead of "RAW", and held in the same folder at the raw data. Metadata will include methodological information, associated calibrations, quality control samples (e.g., certified reference material), and (if applicable) elemental chromatographs of speciated samples. Any changes in processing of raw data will result in a new file with an updated version number, and metadata within that file to indicate changes made. Raw data files are commonly saved in .csv format, but processed data files will be available in .xlsx format (for internal project use; so that graphical and tabulated information can also be stored). Computers housing instrumental data will be periodically backed up (after each analytical run) and files will be uploaded to the MEGA project Box folder (via Penn Box account).
Sequencing data: Metagenomic sequence information will be shared and preserved via uploading to NCBI's Sequence Read Archive as well as to the iMicrobe database, and we will conform to their metadata and formatting requirements.

Data and metadata formats:
Data will be achieved as recommended by the Arctic Data Center Data Format Policy. This policy supports upload of open-source, ubiquitous and easy-to-read data formats. As such, we will upload data as .csv, .txt and .png files when appropriate. Any data held in proprietary formats (e.g., .xlsx) will be converted to the aforementioned open-source formats before submission to the Arctic Data Center. All ORCiD information for individuals collecting and analyzing the data will be provided so that appropriate credit is given.

Data distribution and reuse:
In accordance with NSF Office of Polar Program guidelines, data generated during this project will be disseminated in scientific publications, made freely available on the NSF Arctic Data Center, and made available via email requests within two years of collection.
Data distribution will be managed by the respective instrument users according to the above agreement. Quality control data (standards, drifts, spikes, and reference material) will be available alongside sample data.

Data archiving and preservation:
The project will utilize shared data folders linked to a password restricted University of Pennsylvania managed cloud storage location (Box). This cloud Box folder will be accessible to all project personnel, and is continuously backed up to the cloud.
Data from the project will either (i) be made available with published manuscripts where it will be uploaded to the NSF Arctic Data Center or (ii) uploaded to the NSF Arctic Data Center within two years of collection, as per data and metadata formats above. The choice of (i) or (ii) will be determined by whichever is sooner, and will follow the most up-to-date NSF Office of Polar Programs data distribution guidelines. The Arctic Data Center is freely accessible, and data will be held in perpetuity. The data manager (Hawkings) will ensure all project team personnel follow Arctic Data Center guidelines to provide accurate and complete documentation for data preservation to ensure data will be accessible and available after project funded has finished.

Additional comments:

The data management plan will be reviewed annually by the project team during the annual workshops, with revisions made as appropriate.