DATA MANAGEMENT PLAN

**Metabolomics Mass Spectral Data**:

We acquire metabolomics data on a Waters TQS triple quadrupole LC-MS and a Thermo QExactive (Orbitrap) LC-MS. Management and processing of these data types are similar, but some aspects differ. Each instrument has a connected 8 bay, 16 terabyte RAID Synology box housed in the Ingalls Lab that is maintained by the School of Oceanography IT support team comprised of two staff. All spectral data is temporarily housed on these systems while it is being processed. All raw spectral data is also backed up to Google Drive on a regular basis.

TQS targeted data is processed using Skyline (peak integration) and these data are then quality controlled and normalized using an in house R-based code that is publically available in the Ingalls Lab GitHub repository (https://github.com/IngallsLabUW). The output of this pipeline is a flat file with compound name, retention time, peak area and m/z fragments monitored.

QExactive untargeted data are converted from Thermo's .raw format to the standard mzXML format, prior to peak picking using XCMS or MSDial. These data are filtered, quality controlled, normalized and annotated using an in house R-based pipeline that is available in the Ingalls Lab's GitHub repository.

Our goal is to have an open access archive of our data. File format standards for these data sets and electronic dissemination and preservation plans are as follows. The raw data and instrument settings from each mass-spec run are saved in the manufacturer's native file formats. The files are converted to the open community standard mzXML file format. These data are stored temporarily on our in house Synology drive and are also accessible upon request through our Google Drive. In addition, the metadata and tabular results will be made available in a timely fashion and deposited in publicly available databases. Data are being deposited with Metabolights (http://www.ebi.ac.uk/metabolights/), housed by EMBL-EBI, who will assign our data an accession number that can be cited by the Biological and Chemical Oceanography Data Management Office (BCO-DMO, http://bcodmo.org/resources). Data is also disseminated through publication of peer-reviewed articles.

Data collected under the project will be made available to the public with as few restrictions as possible. For experimental work, we plan for publication of most data with metadata after or in conjunction with primary publication of results, or at most two years after the completion of the study.
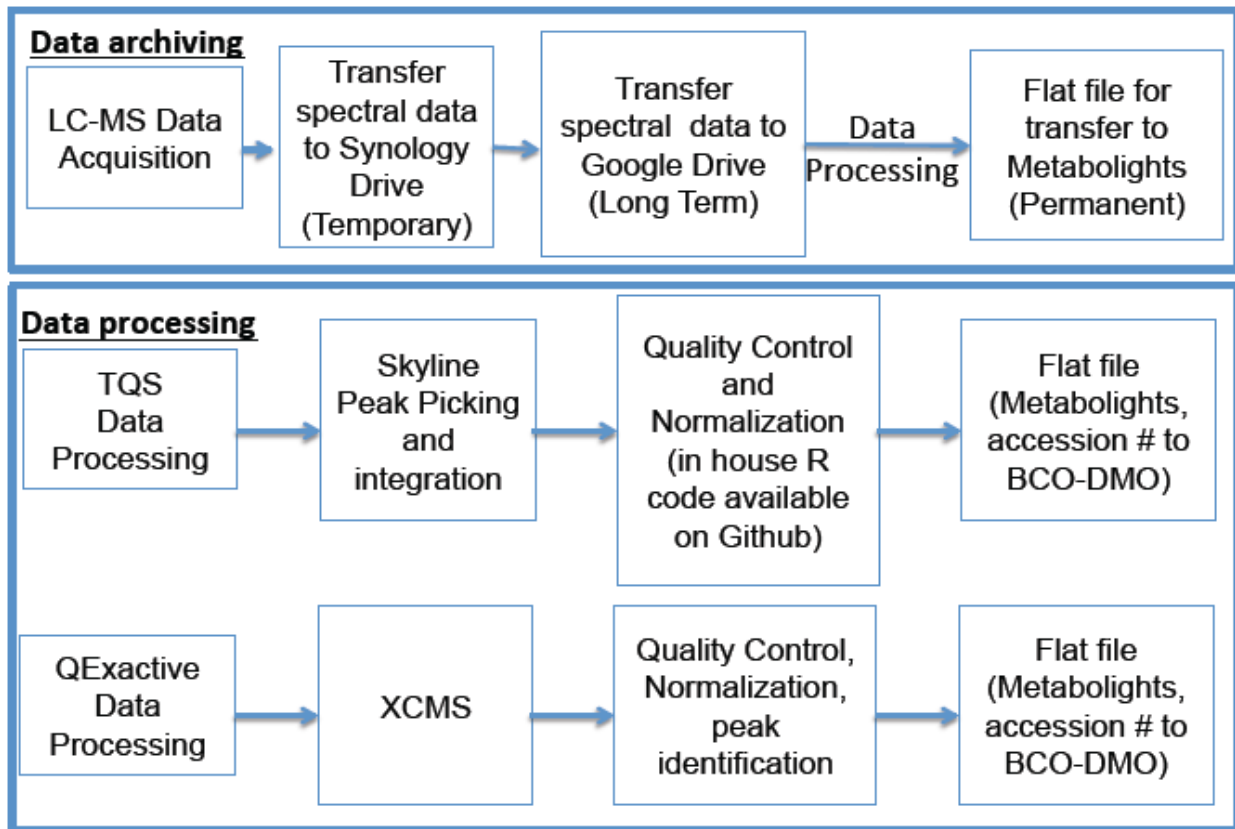
Figure 1: Workflow diagram illustrating how mass spectral data is acquired, processed and shared in the Ingalls lab.