

Data Management Plan – Kujawinski, February 2016

Data types:

The proposed project will generate two types of mass spectral data. First, we will generate chromatographic mass spectral (LC/MS) data collected on the FT-ICR MS which includes the ion chromatograms and the resulting lists of retention times and high resolution mass:charge values for individual ions. These data are processed through R-based scripts and exported to Matlab for further analysis. Fragmentation spectra (MS/MS) are generated for selected ions in these analyses and will be stored on the RAID-V server. Retention times, mass:charge values, MS/MS spectra and putative identifications for each mass spectral feature will be stored in an in-house database (Longnecker et al., 2015). The second type of data is LC/MS data collected on the triple-quadrupole (TSQ) mass spectrometer. These data include extracted ion chromatograms for selected precursor-fragment pairs as a function of retention time along the column. The primary data is processed by MAVEN (Melamud et al., 2010) and the quantity of each compound is exported to Matlab or R. All data will be stored in formats (e.g., spreadsheets, PDF files) that are readily accessible by common software.

Data availability:

All data will be available from the Kujawinski group upon request. We are working with the MetaboLights repository (<http://www.ebi.ac.uk/metabolights/>), which is sponsored by EMBL-EBI in the United Kingdom. Although they do not specialize in the storage of untargeted metabolomics data, they have welcomed our datasets this year and we plan to send them all datasets generated in this project. Datasets are given an accession number, which can be included in publications.

Data archives:

Laboratory notebooks containing primary data will remain in the laboratories of the PIs. Chromatographic primary data and data from other analyses is stored for at least 3 years but lists of retention times and compound abundances can be stored indefinitely due to the reduced file size. The RAID server at the WHOI FT-MS facility is backed up nightly by the WHOI Computer and Information Services department.

Citations

- Longnecker, K., Futrelle, J., Coburn, E., Kido Soule, M.C., and Kujawinski, E.B. (2015). Environmental metabolomics: databases and tools for data analysis. *Marine Chemistry* 177, 366-373. doi: 10.1016/j.marchem.2015.06.012
- Melamud, E., Vastag, L., and Rabinowitz, J.D. (2010). Metabolomic analysis and visualization engine for LC-MS data. *Analytical Chemistry* 82, 9818-9826. doi: