

Data Management Plan

Our project will include several forms of laboratory data, including 1) microscopy, 2) physiological, 3) proteomic, and 4) metabolomic data.

Types of Data

Microscopy: We will collect image data on cell structures using fluorescence and transmission electron microscopy (TEM), as well as routine cell counts to estimate growth and to normalize other measurements.

Physiology: We will generate fluorescence data (FIRe fluorometer), flow cytometry data using software developed by Satlantic (Halifax, CAN) and Accuri (BD, Franklin Lakes, NJ), respectively. We will also generate ^{14}C isotope data that will be counted on a Beckman liquid scintillation counter, as disintegrations per minute (DPM). These data will be exported and processed in Excel.

Proteomics: Proteomic data will include peptide sequences produced by Bioproximity (Chantilly, VA) using LC-MS/MS, and analyzed using four different search algorithms: NCBI's OMSSA, the GPM's X!Tandem, X!Hunter and K-score. All data will be quantified using spectral counting.

Metabolomics: Michigan State University Mass Spectrometry and Metabolomics Core will generate experimental mass spectra on different platforms and in-silico computer generated mass spectra as well as chromatograms. We will include taxonomy, study design, time-course information, standardization and quantification data. Processed spectra, raw chromatograms and annotated biological study results will be publicly shared under open-data licenses. For minor results or smaller updates, data will be directly shared with the public.

Format and standards:

Microscopy: Microscopy data will be collected as raw Zeiss LSM (confocal) or standard TIF (JEOL, TEM). Images can be exported in a variety of data formats (e.g. TIF, JPG).

Physiology: FIRe Fluorometer data is collected using Satlantic software FIReView, FIRePro and FIReCont, all of which run in a DOS OS. Data files (.dat) will be exported and imported into Excel for formatting and further statistical analysis (e.g. with SigmaStat). Flow cytometry data will be acquired using BD Accuri Software, producing data files in FCS 3.0 format, allowing import into other programs such as FCS Express and FlowJo. Data can also be copied in a spreadsheet format from Accuri software, and pasted into an Excel file for further analysis. Scintillation counter data is saved as a tab separated variable (tsv) file, and printed as a hard copy, and will be imported into an excel spreadsheet.

Proteomics: Proteomic data include mass spectrometer binaries in RAW format, peak list files in MGF format, search parameters in XML format and search outputs in XML format.

Metabolomics: All raw mass spectrometry data and method files will be backed up daily to a dedicated server, and archived for long-term storage on magnetic tape at the Michigan State University Research Technology Support Facility (RTSF). Though spreadsheets of integrated peak areas for assorted isotopologs will be transmitted to the PI electronically, the raw data files will also be converted to NetCDF and/or mzXML formats and transmitted to the PI's laboratory if non-routine data processing, which can be performed using various open-source software platforms, is deemed useful.

Data Access and Sharing:

Microscopy: We will also make data available on our website and by request.

Physiology: We will also make data available on our website and by request.

Proteomics: Data will be deposited in a data repository such as Tranche (proteomecommons.org), and will be made available by request. They will also be submitted to EBI's PRIDE repository.

Metabolomics: Protocols and processed mass spectrometry data will be disseminated to the broader community in the form of electronic Supplementary Materials for manuscripts submitted for publication.

Policies for Data Use:

We will generally apply the least restrictive data and program licenses to generate the highest impact for collaborative efforts. Data will be freely available for commercial and non-commercial re-use after publications in peer-reviewed journals with associated funding information.

Plans for Archiving Data/Samples:

Microscopy: Microscopic images will be published, both in print and online as journal articles or as supplemental materials.

Physiology: These data will be published, both in print and online as journal articles or as supplemental materials.

Proteomics: Proteomic data will be deposited in a data repository such as Tranche (proteomecommons.org). We will back up all data on WHOI's data storage system, on CD/DVDs, and on a portable hard drive. Data is also archived at Bioproximity using a fail-safe RAID system.

Metabolomics: Data is immediately stored and copied to fail-safe RAID systems. All data will be additionally mirrored and backed-up at MSU plus an independent backup system stored at WHOI. The data indexing and access system in place is SetupX, a web-based LIMS-system that allows investigators world-wide to detail and set up a biological experiment and download experimental results and raw-data. We constantly research new storage models including P2P storage and cloud-storage.