Solving Critical Problems in Plant Science - from Bench to Market

A Data Management Plan created using DMPTool

Creator: Samuel Hazen

Affiliation: University of Massachusetts Amherst (UMass Amherst)

Template: National Science Foundation (NSF)

Last modified: 02-05-2017

Copyright information:

Solving Critical Problems in Plant Science - from Bench to Market

Data and Materials Produced

Because of the interdisciplinary nature of our proposal, data produced will range from observational and experimental data collected in the lab and field, to large bioinformatics datasets and datasets generated through computer-simulated experiments. These data will be mainly in digital format. Several projects will generate 'omic level data; genomics, RNA and DNA sequencing from our Illumina platforms; proteomic, peptide mass files from mass spectrometry; metabolic, compound mass data from mass spectrometry. Microscopy and biophysical characterization will produce image files.

Standards, Formats and Metadata

Data gathered will typically be in the following formats: MS Excel (.xls), MW Word (.doc), Comma Separated Values (.csv), Portable Document Format (.pdf), Joint Photographic Experts Group (.jpg), Tagged Image File Format (.tiff), sequence (.fasq.gz, fastq), sequence mapping (.bam), sequence alignment mapping format (.sam), peak calling (browser extendable data [.bed]), and mass spectrometry files requiring proprietary software for Orbitrap Thermo (.raw) and Synapt Waters (.raw).

Roles and Responsibilities

This large interdisciplinary project will employ standardized data management program. Each data set will be linked to a project description that describes the purpose of the research, the methods used to generate the data and the experimental design, the period of time data were be collected and if the data will be updated. The Program Director will be responsible for ensuring the implementation of the data management plan with a specific check at each student's annual review. Each researcher will be required to maintain and updated laboratory notebook, either digital or hardcopy. Here we will implement best practices followed industry to assure documentation of the generation of intellectual property. Each week, students will have their notebooks reviewed and notarized by another member of the laboratory.

Dissemination Methods

To facilitate file access and sharing, we will develop a detailed plan for sharing data between collaborators on each project, including the use of secure cloud-based access such as Box, available to University faculty and their collaborators. Generally, participants will be expected to archive and make final datasets publicly available within two years of collection, or as soon as they are published, whichever comes first. Sources mentioned above, such as DMPTool, are available to help to determine each lab's or individual's needs for restricted access (of proprietary or personal information). We will use ScholarWorks as the standard repository for all datasets arising from the program.

Policies for Data Sharing and Public Access

As part of facilitating increasingly complex webs of collaboration, as well as holding members of a collaboration responsible for the data they produce, expectations for project deliverables and plans for disseminating deliverables, when applicable, will be developed at the start of a project and revised as required during the collaboration. Examples of steps collaborators will take to facilitate productive policies for data re-use and re-sharing include:

- Creating a list of participants, by section of a project, for all projects being proposed so that credit can be correctly attributed,
- Including each contributor's expectations for acknowledgement,
- Specifying if data are under license such as common data licenses from Creative Commons or Open Data Commons.

Archiving, Storage and Preservation

Providing a dedicated place for storing data that can be accessed by all collaborators is expected to be outlined *prior to data collection*. Our research group has discussed numerous modes of file sharing including the UMass "Box account," or ScholarWorks, both of which provide storage and local file sharing in conjunction with safe data practices including storage of data in multiple places and cloud based solutions.