

Plan Overview

A Data Management Plan created using DMPTool

Title: Solving Critical Problems in Plant Science - from Bench to Market

Creator: Samuel Hazen

Affiliation: University of Massachusetts Amherst (umass.edu)

Principal Investigator: Samuel Hazen

Data Manager: Samuel Hazen

Funder: National Science Foundation (nsf.gov)

Funding opportunity number: NSF 16-503

Template: NSF-BIO: Biological Sciences (2015-)

Last modified: 02-05-2017

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Solving Critical Problems in Plant Science - from Bench to Market

Data and Materials Produced

Describe the types of data, physical samples or collections, software, curriculum materials, and other materials to be produced in the course of the project. (For collaborative proposals, the DMP must cover all the various data types being collected by each collaborator.)

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

Describe the standards to be used for all the data types anticipated, including data or file format 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

Describe the roles and responsibilities of all parties with respect to the management of the data (including contingency plans for the departure of key personnel from the project).

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

Describe the dissemination methods that will be used to make data and metadata available to others during the period of the award, and any modifications or additional technical information regarding data access after the grant ends.

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

Describe the PI's policies for data sharing, public access and re-use, including re-distribution by others and the production of derivatives. Where appropriate, include provisions for protection of privacy, confidentiality, security, intellectual property rights and other rights.

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

Where relevant, describe plans for archiving data, samples, software, and other research products, and for on-going access to these products through their lifecycle of usefulness to research and education.

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.
