

## Genomics DMP

The proposal will generate research data, models, and software for the analysis of omics data and whole genome models. We will also generate mutant *Arabidopsis* lines.

*Expected Data:* The research in this proposal will produce multiple data types: data files of phenotypes from plant studies, image files in 2-D, 3-D and 4-D, instrument measurement files such as next-generation genomic sequencing data and metabolomics data. We will also be creating new software and algorithms in areas such as image and data processing along with mathematical models of plants and their metabolism.

*Data Format:* The data will be stored as raw data files and in processed form. The files will be stored in the community-accepted formats, e.g., next generation sequencing data will be stored as .bam and count files.

*Access to Data and Data Sharing Practices and Policies:* All raw and processed data along with descriptive metadata will be made publicly available at appropriate public databases (e.g., the Gene Expression Omnibus (GEO) database repository and the Short Read Archive for transcriptome data, and MetaboLights or other metabolomics database for metabolomics data) or on project websites after publications have been accepted.

*Access to Updated Models:* All models and simulation data will be made publicly available as supplementary data in publications and on project websites after publications have been accepted. Scripts and data files for running the models will also be made available.

*Access to Mutant Lines:* All mutant lines will be submitted to the Arabidopsis Biological Resource Center, which allows unlimited access for research purposes. The lines will also be available directly from our laboratories.

*Open Source Software Distribution and Commitment:* We will make any software tools developed under this grant publicly available using platforms such as *github*, which will also help track different program release. We are also firmly committed to using open source software for our tools as much as possible. However, some of the software will be developed in Matlab, an engineering software environment and although our routines may be freely available, the software environment needed to run them must be licensed.

*Archiving of Data:* When appropriate disciplinary public data repositories exist, the data and descriptive metadata will be submitted for long-term public access. The biological models will be submitted to *biomodels.org* for long-term retention. Other data sets will be stored in the ISU data archives for a minimum of five years.