
Plan Overview

A Data Management Plan created using DMPTool

Title: STARS - Soil Testing, Analysis, and Risk-management System: Developing precision molecular tools to detect, predict, and enhance soil health

Creator: Mary Burrows -ORCID: [0000-0003-3765-0405](https://orcid.org/0000-0003-3765-0405)

Affiliation: Montana State University (MSU) (montana.edu)

Funder: United States Department of Agriculture (usda.gov)

Funding opportunity number: USDA-NIFA-SCRI-009081

Template: USDA-NIFA: National Institute of Food and Agriculture

Project abstract:

Specialty crops have direct farm gate value, benefit rural and urban communities, and enhance farm sustainability. Soilborne diseases are a major limitation to yield and quality. We propose to work closely with growers in five states to develop and deploy a testing service called STARS (Soil Testing, Analysis, and Risk-management System) to detect soilborne plant pathogens and measure soil health with a focus on pulse crops including dry pea, lentil, chickpea and dry bean. We will mirror a system developed in Australia which has had substantial success in providing practical information to growers and consultants through applied research and training. We will explore extending the existing paradigm by adding soil health measures. The ability to predict disease prior to planting will benefit specialty crop growers in the short-term, allowing them to make disease mitigation decisions and increase profitability. In the long-term, we will link soil health indicators to pathogen populations and disease intensity, determining the most important measurements of soil health involved in mitigating disease. These measures will then be used to make long-term decisions about land management that will mitigate disease, increase nutrient availability and plant health, reduce excess fertilizer inputs, and decrease the negative environmental consequences of poor soil health. This project is sustainable because testing services will be offered by a collaborating nonprofit lab, and testing tools will be available for growers whose cropping systems face persistent or new pest threats. We will also enhance communication with growers and industry professionals including bankers and processors, support underserved growers including young producers, strengthen our connections with organic growers, and train the next generation of scientists in research, extension, and social science methods.

Start date: 10-12-2022

End date: 09-30-2026

Last modified: 05-13-2022

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

STARS - Soil Testing, Analysis, and Risk-management System: Developing precision molecular tools to detect, predict, and enhance soil health

Expected Data Type

Describe the type of data (e.g. digital, non-digital), how it will be generated, and whether the data are primary or metadata.

- Research examples include: lab work, field work and surveys.
- Education examples include: number of students enrolled/participated, degrees granted, curriculum, and training products.
- Extension examples include: outreach materials, number of stakeholders reached, number of activities, and assessment questionnaires.

Extension outreach summaries will be collected to meet the annual reporting deadline. Data will be collected in an excel spreadsheet. Data on presentations will include presenter name, date, location, title, venue (audience), and the number of participants. We will also record extension outputs such as alerts, publications, videos, podcasts, and social media, including publications including date published, and as required, the number of copies, internet location, and the number of hits/views by the reporting date. USDA-NIFA will be tagged in social media outreach related to this project. This data will be reported in REEport annually. Education data, including the number of students, degree type, date granted, and publications involving student participants, will be reported in REEport annually. Educational opportunities associated with international exchange opportunities will be reported in REEport annually. Administrative data will be collected, including minutes from formal meetings such as PI meetings, stakeholder advisory meetings. Data will be stored in Microsoft Word or EndNote files. Information will be available on request to the PD and Project Manager. Notes will include participant names, meeting dates, agenda, and notes associated with the meeting. Data will be archived in Microsoft Teams. Research data from soil samples, disease assessments and bioassays, soil health indicators, qPCR genetic data, soil microbial data, and weather data will be collected from each grower field, including sampling date, GPS coordinates, grower name, county, state, soil sampling depth, and field history. The GPS coordinates will be generated using a hand-held GPS unit (ex. Garmin eTREX). Disease ratings will be taken using a visual scoring system. The data will be recorded in a handwritten format and then entered into an excel spreadsheet. Data will be stored on computers and the cloud using Microsoft Teams and OneDrive. Each location and subsample will receive a unique indicator coding to develop databases by data collection type. 16S rRNA and ITS amplicon sequence data and microbiome data will be stored on Penn State's high-performance research cloud. Evaluation data from producers and ag consultants, and other stakeholders will be collected using paper surveys and questionnaires, a web-based survey platform (Alchemer or Qualtrics), and audio recordings (when appropriate). All paper-based data gathering will be converted to either tabular form or digital documents. The paper version retained by the evaluation team for the duration of the project and then destroyed. Data will be stored on computers and the cloud using Microsoft Teams and OneDrive. Summary of evaluations will be available annually.

Data Format

For scientific data to be readily accessible and usable it is critical to use an appropriate community-recognized standard and machine readable formats when they exist. If the data will be managed in domain-specific workspaces or submitted to public databases, indicate that their required formats will be followed. Regardless of the format used, the data set must contain enough information to allow independent use (understand, validate and use) of the data.

All evaluation surveys will be cleaned and presented in tabular form, with appropriate metadata either associated with each record or included in metadata sheets within an Excel workbook. For research data, initial data will be collected using Excel workbooks at individual institutions. Data will be analyzed for quality control from which shareable databases using software such as SQLite will be created. Each site-specific database will be maintained with a unique site code identifier. This helps to keep the size and complexity of databases manageable. SQLite also links easily to the R and RStudio statistical programs via the RSQLite package, enabling the development of reproducible and shareable analysis coding for the different methods applied in this proposal. The Esker Laboratory also maintains a GitHub page (@PSUPlantEpidemiology), which will be at the end of the project for sharing code and quality-controlled databases used in analyses after individual researchers have completed their respective analyses for thesis and manuscript writing. Reproducible statistical analyses will be in .R format (R programming language), which links with RStudio's script writing and markdown capabilities. Specific R projects will be created for sharing of entire analyses within the team and to interested persons for training or other analyses. Genomic sequence data will be stored as fastq files, the preferred format for upload onto NCBI's Sequence Read Archive.

Data Storage and Preservation

Data must be stored in a safe environment with adequate measures taken for its long-term preservation. Applicants must describe plans for storing and preserving their data during and after the project and specify the data repositories, if they exist. Databases or data repositories for long-term preservation may be the same that are used to provide Data Sharing and Public Access. Estimate how much data will be preserved and state the planned retention period. Include any strategies, tools, and contingency plans that will be used to avoid data loss, degradation, or damage.

Throughout the duration of the project, all internal data sharing within the team will be done using password-protected cloud-based storage provided by the Montana State University Microsoft enterprise license (Teams and Sharepoint). Any paper records will be retained in locked offices by appropriate team members for the project's duration and retained for at least 3 years after the end of the project. Databases using software like SQLite will be developed and maintained at Montana State University for use by team members and long-term storage. A backup version of all databases to be used for statistical analysis will be created for storage at The Pennsylvania State University using similar software. Primary analyses will be developed at Penn State using RMarkdown and R Project. This enables duplication of coding for team members while maintaining a master version of the coding for long-term preservation and sharing.

Data Sharing and Public Access

Describe your data access and sharing procedures during and after the grant. Name specific repositories and catalogs as appropriate. include a statement, when applicable, of plans to protect confidentiality, personal privacy, proprietary interests, business confidential information, and intellectual property rights. Outline any restrictions such as copyright, confidentiality, patent, appropriate credit, disclaimers, or conditions for use of the data by other parties.

We will use a combination of data access and sharing procedures. During the course of the project, data access and sharing for team members will be through the lead institution (Montana State University) using password-protected cloud-based storage. Statistical analyses and coding will be developed at Penn State and the Soil Health Institute and shared with team members. Individual institutions will retain ownership of local data. Students will be trained on data management and reproducibility and receive training and certificates via their respective institutional review boards. We will also share datasets and links to published manuscripts via the Ag Data Commons. The Esker Laboratory at Penn State houses reproducible code via GitHub, including developing any analytical tools (apps). These software tools can also be shared in the Ag Data

Commons under Software & Tools. The team also works closely with the iPIPE.org team, which will enable long-term storage of tools and models, improving the link with local weather data. The iPIPE.org maintains an approach where each researcher maintains exclusive rights to the data for a three-year period before this would be made available to the broader community. At time of peer reviewed publication 16S, ITS and shotgun metagenome raw sequence data will be uploaded to NCBI's Sequence Read Archive with metadata pertinent to analyses.

Roles and Responsibilities

Who will ensure DMP implementation? This is particularly important for multi-investigator and multi-institutional projects. Provide a contingency plan in case key personnel leave the project. Also, what resources will be needed for the DMP? If funds are needed, have they been added to the budget request and budget narrative? Projects must budget sufficient resources to develop and implement the proposed DMP.

Dr. Mary Burrows, Project Director, Montana State University, will lead and oversee the implementation of the DMP. Dr. Paul Esker, Co-Principal Investigator, Penn State University, will lead the development of analytical and statistical tools used in the project and will lead the training of graduate students on the use of open-source software. He will also ensure that data and code sharing adhere to the methods outline in the DMP. Dr. Sharifa Crandall will lead management of genetic data of soil pathogens and microbial communities. She will ensure DNA sequences are made available to the public on databases (NCBI GenBank). Dr. Elizabeth Rieke will lead management of soil health indicator and inherent soil property data. She will ensure soil health indicator data and associated metadata is shared via the Ag Data Commons at the completion of the project.

Planned Research Outputs

Service - "Soil Extraction Method"

A robust, efficient and accurate large volume soil extraction method will be developed for soilborne plant pathogen detection

Data paper - "Primers for pathogen detection"

Primers for molecular detection of plant pathogens will be developed, tested, and validated on a broad range of soils

Data paper - "Disease thresholds will be established for plant pathogens"

Disease risk thresholds will be developed for target plant pathogens using results from qPCR, bioassays, and in-field disease assessments

Interactive resource - "STARS system"

Risk models will be developed and tested with participating farmers to predict disease risk and implement mitigation strategies.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PI?
Soil Extraction Method	Service	2024-11-01	Open	None specified		Creative Commons Attribution Share Alike 4.0 International	None specified	No	No
Primers for pathogen detection	Data paper	2025-11-01	Restricted	None specified		Creative Commons Attribution 4.0 International	None specified	No	No
Disease thresholds will be established for plant p ...	Data paper	2025-11-01	Restricted	None specified		Creative Commons Attribution 4.0 International	None specified	No	No
STARS system	Interactive resource	2025-11-01	Open	None specified		Creative Commons Attribution 4.0 International	None specified	No	No