

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D1FH1J>

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)

Principal Investigator: Mary Burrows, Frankie Crutcher, Uta McKelvy, Kristal Jones, Audrey Kalil, Zack Bateson, Sharifa Crandall, Elizabeth Rieke, Nonoy Bandillo, Paul Nugent, Tim Seipel, Paulo Flores

Data Manager: Mary Burrows, Frankie Crutcher, Uta McKelvy, Kristal Jones, Audrey Kalil, Zack Bateson, Sharifa Crandall, Elizabeth Rieke, Monica Brelsford, Nonoy Bandillo, Paul Nugent, Tim Seipel, Kevin McPhee, Paulo Flores, Leo Bortolon, Michael Wunsch, Malaika Ebert, Darren Mueller, joe Laforest

Project Administrator: Mary Burrows, Monica Brelsford

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:

Soilborne diseases are a major limitation to yield and quality in specialty crops. In pulse crops such as dry pea and lentil, root rots have led to reduced acres in Canada and the USA. We will work closely with growers in Montana and North Dakota to develop and deploy a framework called STARS (Soil Testing, Analysis, and Risk-management System). Soil and plant sampling will be informed by remote sensing technologies. We will explore soil health measures including the microbiome to determine their predictive value for disease. A descriptive decision support tool called STARS will weight risk factors for root rot. We will investigate the role that environment and herbicide residual have on risk. Solutions will be developed including resistant varieties and seed treatments that can be used on organic acres. Growers were integral to the development of this proposal and are highly engaged. Our long-term goal is to be able to predict the spatiotemporal occurrence of root rot and assist growers to mitigate their disease risk by making informed decisions. Choices made about

land management 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 soil testing services will be offered by a collaborating nonprofit laboratory. We will enhance communication with growers and industry professionals including bankers and processors, support underserved growers including young producers, strengthen connections with organic growers, and train the next generation of scientists.

Start date: 09-30-2023

End date: 08-31-2027

Last modified: 05-15-2023

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 reporting deadlines. Data will be compiled 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/NRS annually.

Education data, including the number of undergraduate and graduate students, degree type, date granted, and publications involving student participants, will be reported in REEport/NRS annually. Training and presentation opportunities will be reported in REEport/NRS 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. Information will be available on request to the PD and Project Manager, and available in Microsoft Teams to the private channel (group). Notes will include participant names, meeting date, agenda, and notes associated with the meeting. Data will be archived in Microsoft Teams (Sharepoint).

Research data from soil samples, disease assessments and bioassays, herbicide concentrations, 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 (e.g., 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 Excel spreadsheets. Excel spreadsheets will use standardized templates stored in read-only directories in Microsoft SharePoint. The standardization will allow data to be read directly from the spreadsheets into MySQL databases that will be later used for analysis. Data will be stored on computers and the cloud using Microsoft Teams (SharePoint). 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 (Crandall PSU).

Evaluation data from producers and ag consultants, and other stakeholders will be collected using paper surveys and a web-based survey platform (Alchemer), and audio recordings of interviews and focus groups (when appropriate). All paper-based data gathering will be converted to either tabular form or digital documents. The paper version will be 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 (SharePoint). Summary of evaluations will be available in digital and print forms annually.

Unmanned aerial systems (UASs) images will be collected from UAS flights. Individual images will be stored in uncompressed or lossless compressed tiff imagery. Each image will contain metadata listing the geolocation, time, optical parameters, spectral bands, and radiometric calibration (as applicable). After flights, data will be collected from the UAS and stored on a desktop computer in each respective lab, Flores at North Dakota State University (NDSU) and Nugent at Montana State University (MSU), and processed into ortho-rectified mosaics using Pix4Mapper. The original images and the ortho-rectified mosaics will be backed up into the NDSU CCAST (Center for Computationally Assisted Science and Technology) or the MSU RCI (Research Cyberinfrastructure) servers managed by each universities respective IT department. Globus data connections will be established between the servers at each university to maintain a unified dataset accessible and stored redundantly at both locations.

Satellite and geospatial data will be collected from online repositories using Google Earth Engine when available or FTP or other direct access methods as needed. These data will be stored in GeoTiffs, shapefiles, or other geospatial data formats. These data will be initially stored on desktop computers in MSU's precision agriculture lab, then will be backed up onto MSU RCI servers and mirrored to NDSU CCAST through a Globus connection.

Regression Models will be produced using RStudio and will be stored as RData; initially, these will be stored on the desktop of the computer used to process the data, then they will be transferred to the SharePoint shared cloud storage space.

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.

Data Format: 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. 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 entire analyses within the team and to interested persons for training or other analyses. The preferred format for genomic sequences, .fastq files, will be curated by the Crandall Lab (PSU) and saved on NCBI's Sequence Read Archive. Imagery from UAS platforms and satellites will be stored as tagged image file format (TIFF) files with appropriate metadata to geolocate the imagery. Ortho mosaic maps and other raster maps will be stored as geographic tiff or GeoTIFF files to enable georeferencing of the data contained in the file. Vector file maps will be stored as open specification geospatial vector files, such as shape files.

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.

Data storage and preservation: Throughout the duration of the project, all internal data sharing within the team will be done using password-protected cloud-based storage with user-specific access control 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 cloud, long-term storage at The Pennsylvania State University using similar software; this enables duplication of coding for team members while maintaining a master version of the coding for long-term preservation

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.

Data sharing, protection, and public access: 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 with user-specific access control. Software for statistical analyses, data management, and modeling will be stored as an open-source project on GitHub and will be maintained and curated by Nugent, MSU. Crandall, PSU and Reike at the Soil Health Institute will contribute data to be stored and shared on this site for public access. The databases for modeling and resulting analysis will be stored on the MSU RCI and backed up to NDSU CCAST, with backups stored on GitHub and SharePoint. Individual institutions will retain ownership of local data. Students and other personnel at the main participating institutions (MSU, NDSU, PSU) will be trained in data storage and management, reproducibility, and software revision control. They will receive training and certificates via their respective institutional review boards. At time of peer reviewed publication, Crandall will ensure that the 16S, ITS microbial community raw sequence data will be uploaded to NCBI's Sequence Read Archive with metadata pertinent to analyses so that the public will have access to these genetic sequences.

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.

Roles and Responsibilities: Dr. Mary Burrows, Project Director, Montana State University, will lead and oversee the implementation of the DMP. Dr. Sharifa Crandall (PSU) will oversee the management of the genetic data of soil and root pathogens for the holistic microbial community analyses and will collaborate with Dr. Bateson (NAGC) to curate these data and the results from the various PCR analyses and bioassays to ensure that DNA sequences are made available to the public (NCBI GenBank). Dr. Elizabeth Rieke will lead the management of soil health indicators and inherent soil property data. She will ensure that soil health indicator data and associated metadata is shared via the Ag Data Commons at the completion of the project. Dr. Paul Nugent at MSU will oversee management of the MySQL database and modeling components of the project. All PIs will use Microsoft Teams to share and store files.

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 PII?
Soil Extraction Method	Service	2024-10-31	Open	None specified		Creative Commons Attribution Share Alike 4.0 International	None specified	No	No
Primers for pathogen detection	Data paper	2025-10-31	Restricted	None specified		Creative Commons Attribution 4.0 International	None specified	No	No
Disease thresholds will be established for plant p ...	Data paper	2025-10-31	Restricted	None specified		Creative Commons Attribution 4.0 International	None specified	No	No
STARS system	Interactive resource	2025-10-31	Open	None specified		Creative Commons Attribution 4.0 International	None specified	No	No