
Development of high-yielding soybean cultivars with diverse genetic background, disease resistance and seed quality

A Data Management Plan created using DMP Assistant

Creator: Istvan Rajcan

Principal Investigator: Istvan Rajcan

Data Manager: Istvan Rajcan

Project Administrator: Istvan Rajcan

Affiliation: University of Guelph

Funder: Ontario Agri-food Alliance

Template: Ontario Agri-Food Innovation Alliance Template

ORCID ID: 0000-0001-5156-2482

Project abstract:

The University of Guelph's soybean breeding programs, based at Guelph has made a significant contribution to Canadian soybean production through the development of more than 150 soybean varieties over the past five decades. The current project aims to address the following objectives:

1. To increase genetic diversity of Canadian soybean germplasm for high yield and seed composition using genetics derived from Canadian x Chinese crosses.
2. To increase diversity and improve the SCN resistance by stacking resistant genes from Hartwig, Peking, PI 567516C, PI88788, and/or Chinese sources.
3. To enhance white mould (WM) resistance in soybean germplasm by identifying and stacking WM resistance genes and improving durable disease resistance to white mould using omics approaches.
4. To develop new varieties with improved resistance to Sudden Death Syndrome (SDS) by introducing diverse genes from exotic germplasm and PI lines.

Identifier: 12520

Start date: 01-09-2023

End date: 31-03-2024

Last modified: 09-11-2023

Grant number / URL: UG-SI-2023-102393

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 customise 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

Development of high-yielding soybean cultivars with diverse genetic background, disease resistance and seed quality

Data collection

Provide an overview of the data that will be generated, collected or acquired to support this project. If data will be acquired from a third party, specify the source.

The data to be generated in this project are field data from soybean variety trials that are conducted as part of the soybean breeding program at the University of Guelph, Guelph Campus, under the PI, Professor Istvan Rajcan.

The collected data will be used by the PI, his team including technicians and graduate students in the program. The data will include raw data and field trial summaries from a variety of field trials across three locations in Southern Ontario (Woodstock, Elora and St. Marys). It will be used to address the objectives in the soybean breeding program and to provide data to partners such as the Ontario Soybean and Canola Committee (OSACC), which is responsible for public variety trials for soybean in Ontario annually. Some of the data generated will be used by graduate students for their thesis projects.

What method(s) of data collection will be employed?

The field data collection is manual for soybean traits such as plant height and lodging, or by harvest combine for traits such as yield and seed moisture. Once the seed is harvested, protein and oil concentration data will be collected on it using the Near-infrared Reflectance machine we have in the lab.

What types of data will be included?

All data will be numeric.

What software or digital formats will be used to collect, manage and analyze the data?

MS Excel will be used to collect the manual field data using tablets.

AGROBASE Generation II will be used for all data analysis.

Provide an indication of the scope of the data?

We run approximated 50 to 60 yield trials in the field annually. The data from those trials are part of this DMP.

Data storage

Estimate the size of data storage that will be required.

Approximately 5 GB.

Where will your data be stored during the collection, collation and analysis phases of the project?

All the data will be deposited on the "PlantAg Shared Drive (G:)" drive, which is located on the server in the Department of Plant Agriculture as described in detail in the plan above. The contact for this data locally will be the PI, Dr. Istvan Rajcan; email: irajcan@uoguelph.ca

All the data will also be backed up in OneDrive in the University of Guelph data repository. We have arranged for help with this aspect of data storage with Wayne Johnston, Librarian at UG.

The data will be stored in both places there indefinitely.

What backup strategy will be employed?

The backup is already mentioned as being on OneDrive.

How will your data files be organized? What file naming conventions will you use? A brief overview or example would be adequate.

As described in the plan above, the data is preservation ready because it is stored in a known directory on the Plant Ag Dept. server and sorted by year of project. Most of the data will be stored in MS Excel and MS Word formats in directories created by year of study. For each year of research, a Table of Contents file in MS Word that helps to identify and locate each file with data for the experiment in question. These

files will also be converted to .txt file format when backing them up onto the OneDrive.

What metadata will be developed for your data? Will there be supplemental documentation prepared to assist with the interpretation and analysis of your data?

Metadata may include a data dictionary to explain data elements, details about units of measure, abbreviations or codes used in the dataset, instrument and protocol information, survey tool details and much more. In a lab setting, much of the content used to describe data is initially collected in a notebook.

Data archiving and preservation

Will you deposit your data in the UG Agri-Environmental Research Data Repository or an external data repository? If you are opting to not archive your data in a repository, where will your data be housed after completion of your project?

All research data, metadata and code that are generated using Alliance funds or resources will be preserved in a non-proprietary file format and available to provide to the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA) upon request.

Discuss any data transformations that will be needed so your data is preserved in appropriate, non-proprietary formats.

No data transformation is anticipated.

If some of your data will not be preserved, how long will you retain it? Will the non-preserved data be destroyed?

All data will be preserved. None will be destroyed.

Sharing and reuse

Will the data that you archive in a data repository be made available for sharing and reuse by other researchers?

Yes. The data will be freely and openly shared through the UG Agri-Environmental Research Data Repository.

Explain which version of your data or subset of your data will be shared.

Final data will be shared, which includes data summaries from all soybean variety trials. No restrictions.

When will your data be available for discovery by other researchers? Will you impose an embargo on publication of your data? If so, please provide details on the duration of the embargo.

Any time. No embargo is planned.

Will you limit who can access your data? If so, who will that be and why are you limiting the data's reuse?

No limits as to who can view and use the data.

Are there specific license terms you will assign to users of your data?

No.

Restrictions/limitations

Are there limitations or constraints on how you manage your data resulting from legal, ethical or intellectual property concerns?

None.

Would your data need to be anonymized or de-identified before being shared with others?

No.

Confidential information

What information do you want to include in your DMP that should not be publicly shared?

None.

Planned Research Outputs

Dataset - "Yield and agronomic data summaries from yield trials. "

All data from the field trials with soybeans will be the research output that will help us make breeding and selection decisions for the project.

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?
Yield and agronomic data summaries from yield tria ...	Dataset	Unspecified	Open	None specified		None specified	None specified	No	No