Draft privacy guidelines and proposal outline to create a Field-Scale Trial Data Repository for data collected by on-farm networks

Thomas F. Morris[1] and Nicolas Tremblay[2]

[1] University of Connecticut, Department of Plant Science and Landscape Architecture, 1376 Storrs Rd, Storrs, CT 06269; [2] Agriculture and Agri-Food Canada, 430 Gouin Boulevard, Saint-Jean-sur-Richelieu, Quebec J3B 3E6 Canada

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada

Abstract. *Implementing better management practices in corn and soybeans that increase profitability and reduce pollution caused by the practices requires large numbers of field-scale, replicated trials. Numerous complex and often unmeasurable interactions among the environment, genetics and management at the field scale require large numbers of trials completed at the field scale in a systematic and uniform manner to enable calculation of probabilities that a practice will be an improvement compared with current practices. Farmers have been completing such trials in large numbers since yield monitors on combines became widely used in the early 2000s. Upwards of 3000 field-scale trials have been completed on corn and 1500 on soybeans in the US, with the likelihood of many more trials completed in other countries with these crops and other crops like wheat. Creating a curated, anonymized data repository for these types of data that is available for use by the agricultural community, especially farmers, farm advisors and scientists, would enable farmers to grow more food at lower cost with much reduced pollution. This paper contains three parts, all in draft form: 1) an outline of a proposal to establish a Field-Scale Trial Data Repository (FSTDR, pronounced faster) for results from replicated field-scale trials harvested by combines with calibrated yield monitors, 2) a set of guidelines for the privacy and security of the farmer data, and 3) a list of the minimum data collection requirements for trials to be included in FSTDR. We welcome edits and comments to improve this paper. We also welcome collaboration with anyone who is interested to join the Working Group entitled "On-Farm Data Sharing" (https://www.rd-alliance.org/groups/farm-data-sharing-ofds-wg) in the Research Data Alliance organization that has the goal of providing a venue to make these types of data available.*

Keywords. *Data repository; field scale-trials; outline of draft proposal; guidelines for privacy and security of farmer data; Research Data Alliance.*

Part 1. Draft Outline of Proposal to Create a Field-Scale Trial Data Repository

Outcomes for proposal

1. Create a curated Field-Scale Trial Data Repository (FSTDR, pronounced Faster) for results of 3000 field-scale replicated strip trials for corn and 1500 for soybean, that are currently in five separate databases, with strict confidentiality for owners of the data and vetted use of the data by farmers, farm advisors and scientists.
2. Create a business model for future expansion of the repository.
3. Demonstrate the benefits of constructing a data repository by creating an interactive, publicly available web site with guidelines for sidedress N rate applications for corn across the Corn Belt, and by publishing two papers: one on the economics of N fertilization in the Corn Belt and one on the economics of fungicide applications on soybeans.

Where proposal will be submitted

Submit proposal to the National Science Foundation, the USDA AFRI program, or a foundation.

Possible sources of existing results to be placed in the FSTDR.

Environmental Defense Funds Adapt Network, Iowa Soybean Association's On-Farm Network, University of Nebraska Lincoln's On-Farm Network, Indiana State Dep of Agriculture's InField Network, South Dakota State University's On-Farm Network, Kansas State University's Farmer Network, and other networks of farmers in the US, Canada and other countries where yield monitors are routinely used to measure yields.

Proposal methods in brief.

1. Describe data available for data repository
2. Describe business model for data repository
3. Use existing data to write 2 papers to show the benefits of data repository
4. Hire programmer to combine data into one database with capability of online access.
5. Design data repository so farmers and farm advisors can easily obtain the results sorted by Technology Extrapolation Domains (TEDs) (Rattalino Edreira et al., 2017; van Warta et al., 2013)
6. Design an interactive web site associated with the FSTDR that uses the extensive N trial results to make estimates of N needed at sidedress based on the amount of rain in the spring up to the time of sidedress with the N rate recommendations expressed in terms of probability of needing N.
7. Design an interactive web site associated with the FSTDR that uses the extensive results from 292 trials to evaluate the fungicide Headline on soybeans to make estimates of the profitability of applying Headline based on the on the amount of rainfall with the results expressed in terms of probability of obtaining a profitable return from application.
8. Outreach program to educate farmers, farm advisors and scientists about existence of data repository, what data are in repository, and how to access and use repository.
9. Outreach program for graduate students about the need for and benefits of data management plans (Ask DataOne at New Mexico State University of New

Mexico to be a collaborator on the proposal). Perform education at ASA meetings.

10. Outreach program for scientists and graduate students on use of non-traditional statistics that would typically be used to analyze the type of large data sets – logistic and Bayesian analyses. Perform education at ASA meetings.

Part 2. Privacy and Security Principles for Data in the Field-Scale Trial Data Repository

The information in this section is based on the American Farm Bureau's Privacy and Security Principles for Data, which is available at: https://www.fb.org/issues/technology/data-privacy/privacy-and-security-principles-for-farm-data.

Ownership:

Farmers own information generated on their farming operations. The information below describes how the data collected about field-scale replicated strip trials and other whole field trials will be used in the Field-Scale Trial Data Repository – FSTDR (pronounced Faster). Farmers submitting data for use and sharing in FSTDR are responsible for ensuring that only the data they own or have permission to use is included in FSTDR.

Collection, Access and Control:

The explicit consent of the farmer by a contract agreement, either signed or digital, will be required to place data into the FSTDR.

Notice:

Farmers who participate in trials that will have the results and field history data from the trials placed in FSTDR will be notified by email that the data will be placed into the FSTDR with the requirement that the farmer send their consent by email or by another method of consent that is verifiable such as certified mail.

Transparency and Consistency:

The purpose of the FSTDR is to provide a sufficient number of evaluations of the diverse management practices of farmers, from nitrogen management in corn to fungicide use on soybeans, to enable analysis of the results in terms of probabilities of economic and environmental success. Farmers can contact the director of the FSTDR at XXX (the director of the FSTDR is unknown at this time, but after establishment of FSTDR this address will be inserted) if they want information about use of their data or have complaints about how FSTDR is handling their data. The FSTDR will disclose data in the data repository only to third parties who agree to the identical terms as FSTDR has with the farmers who have agreed to place their data into the FSTDR. Policies governing the access to and sharing of data in the FSTDR will be posted on the FSTDR

web site, and the policies will be available for free for downloading by anyone. The FSTDR will not change the customer's contract without his or her agreement.

Choice:

After deciding to place their data into the FSTDR Farmers can opt out of the data repository by notifying the director of the FSTDR by email or other verifiable method of notification. The director of FSTDR will have a maximum of 3 business days to send confirmation to farmers about removal of their data from the FSTDR.

Portability:

Only data that has been anonymized will be placed into FSTDR. The GPS coordinates for trials will not be included in the FSTDR. The anonymization process will entail the removal of all information that could be used to identify the location of a trial. Important information for interpretation of results of trials that is normally obtained by use of the latitude and longitude of the trial, such as daily rainfall, minimum and maximum temperatures, soil series, technology extrapolation domains (TEDs), and other information will be included with each trial such that the specific field cannot be identified. The county and state where the trial occurred will be included in the FSTDR if and only if there are a sufficient number of trials to make the data anonymous.

Terms and Definitions:

The FSTDR uses the definitions below for the terms often used about data collection and sharing in the agricultural community.

1. Farm Data are the results collected from field-scale trials completed to evaluate management practices on farmers' fields and the field history information collected about each trial. A list of the minimum data that comprises "Farm Data" is shown at the end of this document.

2. Third party is any person or organization that contracts with FSTDR to analyze data in the FSTDR to improve farmer management practices.

3) Partner is not a term used in relation to the FSTDR, trial results, and field history data collection.

4. Business partner is not a term used in relation to the FSTDR, trial results, and field history data collection.

5. ATP partners is not a term used in relation to the FSTDR, trial results, and field history data collection.

6. Affiliate is not a term used in relation to the FSTDR, trial results, and field history data collection.
7. Data account holder is not a term used in relation to the FSTDR, trial results, and field history data collection.
8. Original customer data is not a term used in relation to the FSTDR, trial results, and field history data collection.

Disclosure, Use and Sale Limitation:

The results of evaluations of management practices and field histories collected about the fields where the evaluations occurred will be sold to third parties who agree to abide by the same terms and contracts FSTDR has with farmers as described in this document. Farmers will not be notified of the sale of the data because the purpose of the FSTDR is to make this type of anonymized data available to farmers, farm advisors and scientists to improve management practices of farmers. The FSTDR will not share or disclose original farm data with a third party in any manner that is inconsistent with the contract with the farmer. The FSTDR will ensure that agreements with third parties will be identical as this agreement between farmers and FSTDR.

Data Retention and Availability:

Each farmer who contributes data to FSTDR will be given a unique identifying number for their data, so they can verify that the data placed in the FSTDR is the data they provided to the FSTDR, and that no field-specific information is include with their data. Data submitted to FSTDR will reside in the repository for at least 20 years or until the data is deemed not valuable for improving management practices of farmers.

Contract Termination:

Farmers can remove their data from FSTDR at any time by requesting removal by email or other verifiable method on notification.

Unlawful or Anti-Competitive Activities:

The data in FSTDR will not be used for unlawful or anti-competitive activities. The data cannot be used to speculate in commodity markets. The data can be used by scientists such as economists to analyze the data with the intent to inform policy decisions.

Liability & Security Safeguards:

Data in FSRDR will be protected with by using the latest security safeguards to prevent loss or unauthorized access, destruction, use, modification or disclosure. Farmers will be notified within 3 business days by email or within 10 days by US Postal Service mail

of any loss or unauthorized access, destruction, use, modification or disclosure of the data.

Part 3. Minimum data to be collected for results of field-scale evaluations and field history information placed in the Field-Scale Trial Data Repository.

The information below about minimum data to be collected for the FSTDR is based on the minimum data required for field-scale trials in the NutrientStar database managed by the Environmental Defense Fund (http://nutrientstar.org/).

List of minimum data to be reported in FSTDR

A. Soil and Landscape

Required information

1. Soils
    a. Slope
    b. Dominant soil series
    c. Drainage class
    d. TED number
2. Tile drainage in trial area
    a. Depth of pipe
    b. Spacing of pipe
3. Soil test results from 3 years or sooner showing: pH, apparent CEC, soil organic matter, extractable K, P, Ca and Mg, and the method of extraction.

B. Cultural Practices & Management

Required information

1. N Fertilizer, past 2 years: rate, form, timing, placement for all applications
2. Manure, past 2 years: type, rate of application
3. Fertilizer other than N made in year of experiment: rate, form, timing, placement
4. Previous 2 crops
5. Corn hybrid
6. Herbicide applications: rate type application method and timing
7. Tillage implements
8. Cover crop: species, planting date, termination date
9. Long-term conservation practice (type, NRCS code if applicable)
10. Herbicides: rate, type, application method and timing

Optional information to enhance evaluations

1. Manure applications: type, yes/no for past 10 years
2. Tillage: depth and dates

C. Trial information

Required information for nitrogen and pesticide trials. Information required for trials evaluating other practices such as seeding rates, tillage practices, cover crops, etc., will be developed.

1. Dimensions of individual plot
2. Dimensions of entire trial
3. Planting date
4. Corn or soybean variety
5. Seeding rate
6. Dates for N applications
7. Placement of N
8. N formulation
9. Rates of N
10. Active ingredient for N stabilizer
11. Concentration of N stabilizer
12. Tested pesticide application rate
13. Time of application for the tested pesticide application
14. Method of tested pesticide application
15. Harvest date
16. Harvest area
17. Harvest method
18. Individual strip yields

D. Environmental

Required information

1. Daily rainfall 6 months before planting to harvest
2. Annual and seasonal rainfall
3. Deviations from normal rainfall – annual and seasonal
4. Minimum and maximum temperatures from 6 months before planting to harvest

Optional information to enhance evaluations

1. Radiation

E. Economic information

Required information

1. Crop price
2. Cost of N fertilizer or pesticide
3. Additional cost of application for nitrogen or pesticide, if any
4. Cost to use N prediction model.

F. Additional optional information that may be collected

Collection can be from each strip, across each treatment, or across the entire trial area.

1. Soil physical properties
    a. Infiltration rate
    b. Aggregate stability
    c. Bulk density
    d. Available water
    e. Water-filled pore-space
2. Biological properties
    a. Respiration by Solvita
    b. Microbial diversity
3. Chemical
    a. Soil test information as noted above in fall before trial season
    b. EC
    c. Total soil N
    d. Soil nitrate and ammonium to depth of 2 feet in 1-foot sections: preplant, when corn between V4 and V6 growth stage, post-harvest.

References

Rattalino Edreira, J.I., S. Mourtzinis, S.P. Conley, A.C. Roth, I.A. Ciampitti, M.A. Licht, H. Kandel, P.M. Kyveryga, L.E. Lindsey, D.S. Mueller, S. L. Naeve, E. Nafziger, J. E. Specht, J. Stanley, M.J. Staton, P. Grassini. 2017. Assessing causes of yield gaps in agricultural areas with diversity in climate and soils. Agricultural and Forest Meteorology 247:170-180.

van Warta, J., L.G.J. van Bussel, J. Wolf, R. Licker, P. Grassini, A. Nelson, H. Boogaard, J. Gerber, N.D. Mueller, L. Claessens, M. K. van Ittersum, K.G. Cassman. 2013. Use of agro-climatic zones to upscale simulated crop yield potential. Field Crops Res. 143:44-55.