Draft Case Statement to Create a Working Group Entitled "On-Farm Data Sharing"

Submitted to

Research Data Alliance

Submitted by

Tom Morris, PhD

Professor

Department of Plant Science and Landscape Architecture

University of Connecticut, Storrs, Connecticut, USA

thomas.morris@uconn.edu

Nicolas Tremblay, PhD

Research Scientist

Saint-Jean-sur-Richelieu Research and Development Centre

Agriculture and Agri-Food Canada

Saint-Jean-sur-Richelieu, Quebec, Canada J3B 3E6

nicolas.tremblay@agr.gc.ca

25 May 2017

Case Statement for On-Farm Data Sharing WG

Charter

Introduction and Rationale

Farmers have capabilities that they have never had before to critically evaluate management practices using field-scale replicated strip trials. Farmers have gained this powerful capability because yield monitors on combines enable accurate measurement of yields. Networks of farmers have been established around the world to exploit the potential of yield monitors to evaluate management practices at the field level. Networks of farmers have become increasingly common because farmers understand the power of evaluating management practices on their fields and across many fields in a similar agroecosystem. Scientists, and then policy makers, can also find value in data coming from a diversity of agroecosystems as previously unknown G x E x M (Genetics × Environment × Management Interactions) (Hatfield and Walthall, 2015) relationships could be derived from contrasted soil, climatic conditions, genotype evaluations, and farming practices.

Collection of results from strip trials across many farmers' fields requires protocols for data stewardship, that is, for data reporting, sharing and archiving. Most farmer networks have developed data stewardship protocols. The protocols, however, vary from network to network, and the protocols are not easily accessible to people outside the networks. Creation of a standardized set of protocols for data stewardship that are publicly available, especially for confidentiality of the data and for sharing of data, would enable the pooling of results from many networks into one secure database. The protocols would be specific to on-farm research performed at a field scale with yields measured by yield monitors. Protocols developed for more general data collection by farmers such as the Thirteen Principles on Data Privacy and Security from the American Farm Bureau Federation, and those developed by the Agricultural Data Coalition will underpin these specific protocols. One big difference in the specific protocols we will create is that our protocols for on-farm research will include minimum data requirements, which other protocols for data stewardship do not include.

Questions to address in the protocols include life cycle, data quality, data infrastructure, formats, standards, protocols, archives, FAIR principles (Wilkerson, 2016), availability, provenance, stewardship, privacy, property rights, laws, confidentiality and governance. Creation of a standardized set of protocols also would promote the formation of new farmer networks and the collection of many more results from on-farm trials, which would greatly increase the value of a secure database. A secure database open to researchers from around the world based on the guidelines to be created by this WG would be an enormously valuable resource for farmers, farm advisors and policy makers.

As a first step, we aim at combining the results of thousands of field-scale replicated trials completed across a diversity of agroecosystems in the US Corn Belt. The Corn Belt covers much

of the 65 million hectares of maize and soybeans planted in the US. This vast dataset would make possible new and previously unavailable analyses to improve productivity, profitability and environmental stewardship. One example of the type of research that could be completed with such a database is from a proposal submitted to the United States Department of Agriculture (USDA) by three farmer networks in the US. The three networks are seeking to develop an interactive, online tool for improved management of nitrogen (N) across the numerous agroecosystems in the Corn Belt of the US. The tool will provide information that farmers need to create locally adapted N recommendations. The tool will have four main components: 1) risk assessment of late-season deficient and excessive maize N status; 2) uncertainty of yield response in individual trials; 3) probability of an economic yield response for different levels of N fertilization, different N timings and fertilizer sources, different cropping systems, and observed rainfall and soil characteristics within fields based on aggregate data; and 4) statistical power analysis to estimate the number of locations and treatment replications needed to detect a specific yield response of interest to a farmer or agronomist.

The online tool will be based on the analysis of archived information from two types of data collected by three farmer networks: 1) 5,420 systematic surveys of the N status of maize fields from Ohio, Indiana, and Iowa across 13 years, and 2) 812 field-scale, replicated N rate on-farm trials in maize fields from Iowa, Ohio, Indiana, Illinois, Michigan, and South Dakota across 12 years.

This will be the first time data from different farmer networks shall be integrated if the proposal to USDA is funded. These data are only a small part of the data that resides in individual databases of farmer networks in the US. Only three of the six networks in the US were cooperators on this USDA proposal. The other networks were hesitant to contribute their data for several reasons, but the main reason was the lack of guidelines about who would have access to the data, for how long, and for what purposes. The data available in these farmer networks are not only the results of N rate trials but contain results from trials about fungicide effectiveness, plant population studies, the effectiveness of N stabilizers, effects of tillage on yield, and many other topics.

One huge advantage of analyzing results from such a large database of fields over many years is that results can be displayed as probabilities. Typical N recommendations for grain crops are made with little to no estimate of the variability in N needs across fields. Because the variability in N needs across fields and years has been shown to be large (Dhital and Raun, 2016), current N recommendations are much less reliable than needed for widespread adoption by farmers.

An example of how results from large numbers of trials can be used to estimate the probability that a maize field will have deficient or excess N, and some of the factors affecting the N status at the field scale is shown in Kyveryga et al. (2013). This data set contained 56 field-scale,

replicated, two-treatment studies over 2 years in the state of Iowa where the N fertilizer rate was decreased by 56 kg ha^{-1} compared with the rate normally applied by farmers who participated in the trials. The intent of this study was to help farmers decide whether they could profitably reduce their N rates by 56 kg ha^{-1}. The results showed that the probability of increased profit with reduced N fertilizer was reduced by 35% when high amounts of rainfall occurred in June, but was increased by 20% when soil organic matter was high.

Other important advantages of combining data from farmer networks are that meta-analysis techniques are not needed because the data are the raw results from individual strips in the trials. Results from individual strips are preferable to aggregate data because more comprehensive data analyses can be performed to fully understand the treatment effect (Jones et al., 2009). Combined data also are of much greater value to other scientists such as economists who analyze data using different techniques and hypotheses than agronomists. And field-scale trials allow measurement of the effect of spatial variability within fields on yield and profit, which small research plot studies are not capable of measuring. Because management practices by farmers are greatly affected by spatial variation of soil properties (including topography) within fields, field-scale trials are needed to measure these effects.

Deliverables and Outcomes

The deliverables for the On-Farm Data Sharing WG will be:

1. Minimum data requirements for field-scale, replicated strip trials completed by farmers using GPS-guided equipment including combines with calibrated yield monitors.
2. Guidelines for collecting, handling, storage and formatting results and metadata from field-scale, replicated strip trials
3. Guidelines for stewardship of data collected from field-scale, replicated trials completed on production grain fields, which will include guidelines for:
    a. Data accessibility
    a. Licensing options for allowable uses of the data (data sharing)
    b. Curation of the data
    c. Maintaining confidentially of the data

The outcomes for the On-Farm Data Sharing WG will be:

1. Agreement among interested farmer networks in the world to place their data in one common database using the guidelines developed as part of the deliverables. The data and meta data managed by the 6 major farmer networks in the US[1] will be the first data to populate the database with data from other networks in the US and in other parts of the world, especially from countries with many combines having yield monitors such as Canada, countries in western Europe, Australia, and Argentina who are interested to participate, added later.

2. Submission of a proposal to the U.S. National Science Foundation or other funding organization for funding to clean and collate the data that will be entered into the database, and to create a common, secure database for the results of trials and for the field metadata.

When these outcomes are achieved, the guidelines established will serve as a baseline for other networks representing other agroecosystems to follow suit with their own adapted sets of requirements.

[1] **The 6 major farmer networks in the US are**:

1. On-Farm Network managed by the Iowa Soybean Association
2. Adapt Network managed by Environmental Defense Fund
3. Infield Advantage managed by the Indiana State Department of Agriculture
4. On-Farm Research Network managed by the University of Nebraska Extension
5. New York New York On-Farm Research Partnership managed by Cornell University
6. K-State On-Farm Network managed by Kansas State University, Kansas State Research and Extension.

Value Proposition

Society will be the largest benefactor from implementation of the On-Farm Data Sharing WG outcomes. Grain crops will be grown with lower costs and less pollution. Specific benefactors will be farmers and farm advisors, scientists and policy makers.

The tangible benefits for each group are:

1. Farmers and farm advisors will obtain more reliable and accurate recommendations for many management practices that are difficult or impossible to evaluate in small-plot research. Examples of management practices that are best evaluated on a field scale include: fertility management, especially N management; pest management; plant population management and interactions with fertility; soil and fertilizer enhancement products such as N stabilizers and products derived from humic acids, etc. Economic analysis of changes in management practices will also be more accurate and realistic with results from field-scale trials.
2. Scientists will have access to reliable, replicated research results about the effects of changes in management practices on profit and the environment at an unprecedented scale, both geographically and numerically. Given the complexity and diversity of biological and physical conditions in agriculture fields, and the interactions of these conditions with the enormous number of management practices (types and degrees) used by farmers, large data sets of replicated, field-scale trials are needed to categorize practices into probabilities of success. Current methods of research are inadequate to create probability distributions of management practices by environment. The guidelines developed by this WG will enable scientists to publish much more reliable and accurate estimates of which management practices are best used in any environment. Also, a dataset of this quality and magnitude will be used to apply data mining algorithms and machine learning leading to further discoveries of potentially applicable decision rules.
3. Policy makers will benefit by having access to more reliable conclusions about the effect of management practices on profit and the environment. This will enable policy makers to create better informed and effective programs for food production.

Engagement

There are state and regional efforts to create databases of results of replicated field-scale trials to improve recommendations for management practices. The Iowa Soybean Association is the leader in this type of effort in the state of Iowa. The Indiana State Department of Agriculture's INField Advantage program is modeled after the Iowa Soybean Association's program, and the Environmental Defense Fund's On-Farm Network is a similar program except their program crosses state borders to include trials from Ohio, Indiana, Michigan and Illinois. Members of

these organizations are part of the International Society of Precision Agriculture's Community entitled "On-Farm Data Sharing", and representatives of these organizations will be part of the IGAD On-Farm Data Sharing WG.

The 4R Research Fund works to create databases of existing research on nutrient management, and to create new research to increase the size of the databases they are creating. This program is organized and run by the International Plant Nutrition Institute (IPNI). A scientist from IPNI is a member of the On-Farm Data Sharing Community and will be a member of the On-Farm Data Sharing WG.

One goal of the On-Farm Data Sharing WG will be to seek scientists from around the world who are working with farmers, either informally or formally in organizations, to implement replicated field-scale trials harvested by combines for the purpose of improving management practices.

Work Plan

A specific and detailed description of how the WG will operate including:

1. The final deliverables of the On-Farm Data Sharing WG will consist of:
    a. Guidelines for minimum data requirements for field-scale, replicated strip trials completed by farmers using GPS-guided equipment including combines with calibrated yield monitors.
    b. Guidelines for collecting, handling, storage and formatting results and metadata from field-scale, replicated strip trials
    c. Guidelines for stewardship of data collected from field-scale, replicated trials completed on production grain fields, which will include guidelines for:
        i. Who has access to the data
        ii. Allowable uses of the data
        iii. Curation of the data
        iv. Maintaining confidentially of the data
2. Milestones for the WG include:
    a. September 2017. Acceptance of the WG Case Statement by the Research Data Alliance
    b. December 2017. Guidelines for minimum data requirements completed.
    c. February 2018. Guidelines for collecting, handling, storage and formatting results and metadata completed.
    d. April 2018. Guidelines for stewardship of data collected from field-scale, replicated trials completed.
    e. June 2018. Proposal submitted to a funder such as the U.S. National Science Foundation for funding to clean existing data, format data, and create a secure database for placement of data from 6 existing farmer networks.

  f. June 2018. Presentation of guidelines at the International Conference on Precision Agriculture.

  g. July 2019. Presentation of guidelines at the European Conference on Precision Agriculture.

As the bulk of the contributors are coming from the crop science sector, it is unlikely that many of them will be attending the RDA plenaries. Our intention is to seize opportunities stemming from agronomical scientific meetings (for instance, before the launch of the OFDS-WG activities, a poster will be presented at the 11th European Conference on Precision Agriculture (ECPA 2017, July $16-20$, 2017, Edinburgh, UK), and to work in a collaborative environment such as DropBox, which allows for fluid comment and changes to documents. It is expected that contributors will be meeting in person at the many conferences on crop science, agronomy or precision agriculture that are being held several times per year. Example of those are:

- 7th Asian-Australasian Conference on Precision Agriculture (October $15-20$, 2017, Hamilton, New Zealand)
- American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America 2017 International Annual Meeting (October $22-25$, 2017, Tampa, Florida)
- 14th International Conference on Precision Agriculture (June $24-27$, 2018, Montreal, Canada)

• A description of how the WG plans to develop consensus, address conflicts, stay on track and within scope, and move forward during operation.

  Consensus will be built by submitting all drafts of the guidelines and proposals to all members of the WG, and by providing sufficient time, usually 3 weeks, for review of the documents. Conflicts will be addressed by discussion and by building consensus through discussion and email exchanges. With members of the WG located distant from each other discussions will occur by using Skype. Face-to-face meetings will be held at RDA plenaries and at meetings such as the American Society of Agronomy's annual meeting to build consensus. To stay on track and within the scope of the work plan, monthly email exchanges will occur to check on progress of writing the guidelines and proposals.

• A description of the WG's planned approach to broader community engagement and participation.

  We will attend many conferences on crop science, agronomy and precision agriculture, and we will inform the agronomy community about the importance and status of ongoing and completed work within the WG.

Adoption Plan

Agreement among the 6 major farmer networks in the US to place their data in one common database by December 2018 using the guidelines developed as part of the deliverables has been established as an objective of the WG. Also, the submission of a proposal to the U.S. National Science Foundation or other funding organization for funding to clean and collate the data in each of the 6 major farmer networks in the US, and to create a common, secure database for the results of trials and for the field metadata is part of the plan for adoption or implementation of the WG outcomes within the organizations and institutions represented by WG members, as well as plans for adoption more broadly within the community.

Initial Membership

Initial leadership:

Tom Morris	U Connecticut	Thomas.Morris@uconn.edu
Nicolas Tremblay	Agriculture Agri-Food Canada	Nicolas.Tremblay@agr.gc.ca

Initial members (TBC)

Bertin, Patricia	Embrapa, Brazil	patricia.bertin@embrapa.br
Bonnet, Pascal	CIRAD, France	pascalbonnet@cirad.fr
Ciampitti, Ignacio	K-State U	ciampitti@ksu.edu
Clay, David	South Dakota State U	david.clay@sdstate.edu
Craker, Ben	AGCO	ben.craker@AGCOcorp.com
Ekpe, Sonigitu A.Nigeria		sonigitu.ekpe@graduateinstitute.ch
Ferreyra, R. Andres	Ag Connections LLC	andres.ferreyra@agconnections.com
Gullotta, Gaia	Bioversity International, Italy	
Hatfield, Gary	South Dakota State U	gary.hatfield@sdstate.edu
Kyveryga, Peter	Iowa Soybean Association	pkyveryga@iasoybeans.com
Murrell, Scott	IPNI	smurrell@ipni.net
Neveu, Pascal	INRA	Pascal.Neveu@inra.fr
Rabe, Nicole	Ontario Ministry of Ag	Nicole.rabe@ontario.ca

Reverte, Carmen	IRTA, Spain	carme.reverte@irta.cat
Soonho, Kim	International Food Policy RI	soonho.kim@cgiar.org
Stavrataki, Maritina	Agroknow, Greece	maritinastavrataki@agroknow.com
Stelford, Mark	Premier Crop	mstelford@premiercrop.com
Thompson, Laura	U Nebraska-Lincoln	Laura.thompson@unl.edu
Yost, Matt	ARS − U Missouri	Matt.Yost@ARS.USDA.GOV

References

Dhital, S., and W. R. Raun. 2016. Variability in optimum nitrogen rates for maize. Agronomy J. 108: 2165-2173.

Hatfield, J. L., and C. L. Walthall. 2015. Meeting global food needs: Realizing the potential via genetics $\times$ environment $\times$ management interactions. Agronomy J. 107: 1215-1226.

Jones, A.P., R. D. Riley, P. R. Williamson, A. Whitehead. 2009. Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. Clinical Trials 6: 16−27.

Kyveryga, P. M., P. C. Caragea, M. S. Kaiser and T. M. Blackmer. 2013. Predicting risk from reducing nitrogen fertilization using hierarchical models and on-farm data. Agronomy J. 105: 85-94.

Wilkerson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, B. Mons et al. 2016. The FAIR guiding principles for scientific data management and stewardship. Nature. SCIENTIFIC DATA | 3:160018 | DOI: 10.1038/sdata.2016.18 1.