THE INTERNATIONAL SOCIETY OF PRECISION AGRICULTURE PRESENTS THE
13th INTERNATIONAL CONFERENCE ON
PRECISION AGRICULTURE
July 31-August 4, 2016 • St. Louis, Missouri USA

Rationale for and Benefits of a Community for On-Farm Data Sharing

Thomas F. Morris and Nicolas Tremblay

University of Connecticut Department of Plant Science and Landscape Architecture, Storrs, CT 06269; Agriculture and Agri-Food Canada, Saint-Jean-sur-Richelieu, Quebec, Canada J3B 3E6.

A paper from the Proceedings of the
13th International Conference on Precision Agriculture
July 31 – August 4, 2016
St. Louis, Missouri, USA

Abstract.

Most data sets for evaluating crop production practices have too few locations and years to create reliable probabilities from predictive analytical analyses for the success of the practices. Yield monitors on combines have the potential to enable networks of farmers in collaboration with scientists and farm advisors to collect sufficient data for calculation of more reliable guidelines for crop production showing the probabilities that new or existing practices will improve the efficiency of food production. The creation of sufficiently large data sets requires the pooling of data from numerous farmer networks, but such pooling of data is currently not possible because there are no standards for sharing of data across networks. The objectives of this paper are to: 1) provide a rationale for a community for on-farm data sharing; 2) describe the challenges of sharing data from on-farm networks and of sharing research data in general; and 3) identify the benefits of data sharing by reviewing what could be gained if data were shared across existing networks in the Corn Belt of the US. Writing and publishing standards for stewardship of data from farmer networks that would include standards for sharing and confidentiality of the data will encourage the creation of large data bases of results from replicated strip trials. The benefits from large data bases of such results are enormous. The greatest benefit is agronomists would be able to move away from the common practice of analysis that answers only the question whether there was a treatment effect to analyses that provide reliable probabilities of the chances a crop production practice will improve the efficiency of food production, and the magnitude of the treatment effect.

Keywords. *Farmer networks, data sharing, large data sets, replicated strip trials, benefits of analysis of large data sets.*

Rationale for and Benefits of a Community for On-Farm Data Sharing

Introduction

Yield monitors provide farmers with a revolutionary technology to rapidly increase efficiencies in crop production practices. Yields can now be measured easily, accurately and inexpensively at the field scale, and because yields are the primary yardsticks for assessing crop practice practices, farmers can assess practices on a routine basis. However, farmers working alone cannot generate sufficient data to increase efficiencies in crop production practices. That is because crop production practices are greatly affected by environmental conditions, and large amounts of data are required to accurately describe the effects of environmental conditions on crop production practices.

Farmers are cooperating in networks with scientists, crop consultants, agency personnel, and commodity organizations to realize the great potential of yield monitors to improve crop production practices. Farmer networks in the past were often named "farmer-to-farmer networks" (Anonymous 1996), but these types of networks were typically created for learning through discussion while the farmer networks discussed here are created specifically to establish on-farm strip trials for the purpose of learning how to improve crop production practices through discussion of results. A definition of a farmer network is "a group of farmers working with one or more advisors dedicated to learning how to improve farm practices through collaborative, scientific evaluation of those practices, and sharing the evaluation results through meetings in groups or one on one. Farmer networks conduct on-farm research through replicated strip trials using production-scale equipment, and by testing a change in practice or management from the farmer's normal practice following a standard protocol" (Chapman, et al., 2016).

The Indiana Department of Agriculture's INfield advantage program and the Iowa Soybean Association's On-Farm Network are two examples of active farmer networks. Networks enable completion of replicated strip trials on a field scale, and quick analysis and summary of the strip trial results for greater learning from the results by farmers. The full potential of yield monitors, however, is being impeded by a lack of clear guidelines for data stewardship, which is the managing, aggregating, sharing, and accessing of the huge amount of trial results and meta-data collected by farmer networks.

Guidelines for data stewardship would greatly increase the value of the results of strip trials. Guidelines would facilitate combining strip trial results across farmer networks, which would allow creation of large collections of strip trial results with the associated meta-data. Large collections of results with the meta-data are needed to enable decisions about crop production practices based primarily on large databases and not primarily on limited data sets that cannot provide adequate solutions to crop production practices, which results in the decisions often being made based on limited information, experience, unreplicated trials, and/or expert opinion.

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 13th International Conference on Precision Agriculture. EXAMPLE: Lastname, A. B. & Coauthor, C. D. (2016). Title of paper. In Proceedings of the 13th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.

Basing decisions about crop production practices primarily on scientifically robust data is needed for two reasons. First, society is demanding better solutions to problems like excess nitrogen and phosphorus in water bodies caused by agricultural production; and second, our current methods of research are inadequate for creating solutions to problems like excess nitrogen and phosphorus in water bodies because there are no guidelines for sharing and accumulating results from large numbers of trials evaluating productions practices. These two reasons are linked. Without large collections of data with thousands of trials across many years, improvements in crop production practices are difficult to create.

Large collections of results from field-scale replicated strip trials are required to create solutions to problems caused by crop production because crop production is fundamentally a plant growth process. Plant growth processes are biological processes driven by interactions among plants, soils, and environmental conditions that are difficult to predict. Due to the complexity of these interactions, many more trials than typically available are needed to describe the probability that a new crop production practice will be superior to an existing practice. The only way to obtain the amount of data needed across years and fields is by cooperation among farmer networks to create large collections of results from their trials with the associated meta-data.

The objectives of this paper are to: 1) provide a rationale for a community for on-farm data sharing; 2) describe the challenges of sharing data from on-farm networks and of sharing research data in general; and 3) identify the benefits of data sharing by reviewing what could be gained if data were shared across existing networks in the Corn Belt of the US.

Challenges to data stewardship

There are many challenges to data stewardship. To overcome these challenges, the benefits of data stewardship, especially the sharing of data, need to be great enough to make the challenges inconsequential. We consider the benefits to sharing of on-farm data to be enormous collectively for farmers, the agricultural community and society. Benefits to individual researchers, farmers, agricultural consultants, agency and industry personnel are more difficult to show because individuals often are not rewarded for sharing on-farm data. A lack of reward for individuals is only one of many challenges to sharing data from on-farm research, and many of the challenges are not trivial (Tenopir et al., 2011). Below is a summary of the challenges to data stewardship of on-farm data and of research data in general. Data sharing is one component of data stewardship. After reviewing the challenges, we will examine the benefits of sharing results and meta-data from on-farm strip trials.

There are seven primary challenges to data stewardship:

1. Standards for data management plans (DMPs)
2. Ownership of data
3. Confidentiality of data
4. Credit for data shared
5. Fear of misuse of data
6. Tenure process
7. Cost of sharing data.

Standards for data management plans.

Data management is a common challenge for the stewardship of data collected by all researchers whether the data is collected from farms in farmer networks, by graduate students from agricultural research stations, or by any researcher from any disciple (Jahnke and Asher, 2014). The management of data has not typically been taught in graduate school and most graduate students develop their own unique methods for managing data from their research trials. This makes it extremely difficult to combine data from experiments completed by different graduate students. The long-term solution to this problem across all agricultural research is the training of graduate students in the art and science of data management.

Librarians and information management specialists are great resources for learning how to manage data. Libraries at major universities offer instruction and support for creating data management plans (DMPs) (Cross, 2016), and many of those same universities offer use of the DMPTool, which is a commonly used template for data management (https://dmptool.org/). Developing DMPs has only recently become part of performing research. The National Science Foundation began requiring DMPs as part of grant proposals only in 2011 (Metcalf, 2015). DMPs as described by DataOne (Anonymous, 2016) are written plans about how data will be authored and how the data will be managed and made accessible throughout its lifetime. The contents of a data management plan should include:

• the types of data to be authored;

• the standards that would be applied, for example format and metadata content;

• provisions for archiving and preservation;

• access policies and provisions; and

• plans for eventual transition or termination of the data collection in the long-term future".

Without a data management plan, access to data is extremely difficult and combining data from many data sets that do not have a common format is expensive. The NSF does not have a template for DMPs because each science discipline has their own definition of what comprises data. That is another obstacle to sharing data: the need for each science discipline to develop standards for reporting their specific data. With some publishers like the Public Library of Science (PLOS) requiring authors to "make all data underlying the findings described in their manuscript fully available without restriction, with rare exception" (Anonymous, 2016a), data management plans should slowly become a required part of graduate education and grant proposals.

Farmer networks have developed standards for data management and analysis of data, but our experience is that few networks publish their standards. The only network that has published a document that provides some guidance for data management is the On-Farm Network at the Iowa Soybean Association. Their "Guide to On-Farm Replicated Strip Trials" (Kyveryga, et al., 2015) contains a wealth of information about setting up and running on-farm trials, and some information about how to manage data, but the information is not a complete guide to management of data from farmer networks.

Ownership of data

Scientists frequently view data collected in their research programs as owned by the scientists (Tenopir et al., 2011). Most scientists, until recently, imbibed this concept from their major professors. Due to the enormous time, energy and money required to run a research program, and the benefits to scientists of novel and original research results, one would expect scientists to think of results from their programs as belonging to them.

A few scientific disciplines, however, have a long history of sharing data. The field of astronomy has long shared data with the sharing of data from the Hubble telescope and the Sloan Sky Survey as two recent examples (Puniewska, 2014). In the genomics area, a long history of sharing data was codified by the National Institute of Health in 2014 (Anonymous, 2016b) with a comprehensive policy that promotes sharing of genomic data collected in NIH-funded projects. Overcoming the cultural norm of individualistic owning of data by scientists in agricultural disciplines will require time, education, and a secure, easy to use and inexpensive infrastructure for sharing data. Probably the quickest way to increase sharing of agricultural data would be for major grant funders like the National Institute of Food and Agriculture to develop a policy like NIH's to promote data sharing from NIFA-funded projects.

Farmers also view the data collected from trials run on their farms as their data that should not be provided to others without their consent. Farmers who participate in on-farm networks typically allow the results from trials on their fields to be aggregated with other participants to make the results more useful. The On-Farm Network of the Iowa Soybean Association both aggregates and publicly displays results from trials with meta-data about the fields and practices, but only if it is impossible to identify an individual farmer. The On-Farm database of results from field-scale replicated strip trials is probably the most extensive in the US and maybe the world. An example of the type of data publicity displayed by the On-Farm Network is available at this link: http://www.isafarmnet.com/onlinedb/index.php. Other networks also publicly display results from trials for the benefit of the agricultural community. See INfield Advantage web site at: http://www.infieldadvantage.org/results-map/, and the Nebraska On-Farm Research Network at: http://cropwatch.unl.edu/farmresearch.

Combining the results from numerous networks like these across an agricultural region like the Corn Belt in the US would greatly increase the value of the data. The first step to combine the data would be to develop a standardized set of guidelines for stewardship of the data, which would include confidentiality of the data and is discussed below. One of the objectives of the On-Farm Data Sharing Community is to develop guidelines for data stewardship of data collected in farmer networks. The benefits of combining or pooling data across numerous farmer networks will be discussed later in this paper.

Confidentiality of data

Confidentiality of data in agriculture in the past has been more of a concern of industry to protect information related to patents or proprietary products. Most research in the past was conducted on public or private research stations and information about farmers' practices was not part of the reported information. With the ubiquitous availability of yield monitors on combines much applied research has moved to farm fields, and these data sets have become huge in the past 10 to 12 years with the development of farmer networks dedicated to generating results from replicated strip trials. These data sets contain sensitive information including locations of farm fields and usually the name of the farmer along with the yields and practices performed on the fields.

The lack of standards for protecting the confidentiality of this type of information impedes the development of more farmer networks, and makes it difficult to pool data across networks. Expansion of networks and the combining of data sets is important to make the best use of results from replicated strip trials performed by farmer networks. Most of the strip trials are established to answer important applied questions in production agriculture. Due to the effect of environmental conditions, practices and soils on the results of applied research trials, and the interactions among these factors, results of hundreds and most likely thousands of trials across areas like the Corn Belt are needed to obtain reliable answers for the many questions farmers, the agricultural community and society are asking about crop production practices.

Precision agriculture technologies, of which yield monitors are one component, offer great promise to improve crop production practices. But these technologies generate large amounts of farm-specific information, which has farmers concerned about confidentiality of this information. The concerns of farmers about the ownership and confidentiality of data from precision agriculture technologies is similar to the concerns about confidentiality of trial results and meta-data of farmers in farmer networks. The concerns about data from precision agriculture technologies prompted the Farm Bureau to work with a coalition of farm organizations and agriculture technology providers to write and publish an agreement in 2014 about how such data should be handled. Thirty-seven companies and associations have signed the agreement as of 3 March 2016 (Anonymous, 2016c). The document provides an excellent template for standards the On-Farm Data Sharing Community plans to write and publish.

Credit for Sharing Data

Credit for sharing data or reuse of data by others is a challenging problem for scientists (Roche et al., 2014). Scientists rightly want credit for their original thoughts and diligent labor creating new data and prefer to not share their data (Savage and Vickers, 2009), especially in a publicly available data base. Because many benefits accrue from publicly sharing data, scientists desire to protect their data and maintain exclusive use of the data in the future creates conflict between individual scientists and society. Unfortunately, mechanisms to reduce or eliminate this conflict are in their infancy (Roche et al., 2014). More needs to be done to codify guidelines for citation of shared data, and to include citations for data sent to data repositories in the evaluation and promotion of scientists. A great review of both the benefits and obstacles to public data sharing is in Roche et al (2014).

Receiving credit for sharing data is not a top priority for farmers who join farmer networks. Most farmers who join a network are primarily interested in learning how to farm more efficiently. Not receiving credit for sharing data does not create a hardship for farmers or diminish their career as a farmer as it can for scientists.

Fear of Misuse of Data

Fear of misuse of publicity shared data and fear of discovery by other researchers of mistakes in analysis by the original author is a fear for many scientists (Tenopir et al., 2015). Although this fear has been considered secondary to the fear of losing the ability by authors to have exclusive reuse of their data in the future (Roche et al., 2014), it is still a major obstacle to data sharing. Overcoming these fears will probably occur slowly as the scientific culture shifts from data being considered the private domain of individual scientists to one of a collaborative venture among scientists for the benefit of society.

Farmers' fear of misuse of data from their fields usually centers on a breach in the confidentiality of a database, and subsequent use of the data to implicate them in pollution of nearby waterways or non-compliance with a mandated nutrient management plan. The solution to this potential problem is tight security of the database to protect from unauthorized entry, and the training of users of the database to create passwords that are impossible to crack.

The Tenure Process

The tenure process inhibits the sharing of data by making the evaluation of scientists' achievements based almost exclusively on publications and grants the scientist can attribute to themselves (Fecher et al., 2015). These standards for tenure are in opposition to the desire of many grant programs and university administrators to support multi-disciplinary and multi-institution research. Separating an individual researcher's contributions to a project, paper or grant is difficult when there are many contributors. Many young faculty seek out grants that enable clear paths to short-term projects where they can publish quickly and be the primary author on the paper or grant. The tenure process also inhibits sharing data within or across farmer networks because of the short timeframe for tenure, usually only six years, while the reason for sharing date is to enable analyses that include many years and locations for increased reliability of conclusions, which would not leave sufficient time for a scientist on a tenure track to publish a paper about the shared data before the end of the tenure evaluation period.

Cost of Sharing Data

There are two main costs of sharing data. Creating and implementing DMPs is a cost common to all researchers that in the past was not included in the cost of completing the research. This cost will not be a major barrier to sharing data in the future when all research is performed with a DMP and all researchers have similar costs for DMPs.

The second cost of sharing data is the cost to place data in a data repository. For data published in scientific journals that require data sharing, such as *Ecological Monographs* published by the Ecological Society of America, the cost to place data in a digital repository like Dryad, which is frequently used by authors who publish in *Ecological Monographs*, is only $120 (Anonymous, 2016d).

Sharing on-farm data from farmer networks will be much more expensive than sharing data from published scientific articles. There will be a number of costs that are impossible to quantify at this time. A list of the major items with undefined costs would include:

1. Creation of standards for data stewardship
2. Time to gain the confidence of farmers in numerous farmer networks to allow sharing
3. Harmonization of data in different formats
4. Creation of a data repository
5. Curation of the data in the repository
6. Long-term maintenance of the data repository.

Funding for the costs to establish a repository for data from farmer networks most likely will have to come from grants. A part of that grant will have to include a plan for long-term maintenance of the repository.

Benefits to Data Sharing

The potential benefits to data sharing are enormous. Sharing of results from field-scale trials on production agricultural fields has the greatest potential for rapid improvement of crop production practices. Results from farmer networks are most representative of how well a new crop production practice will perform under the varied environments, soils and farmer managements that occur where the trials were conducted. Results from small-plot trials cannot provide as reliable results as field-scale trials because of the effect of many factors but especially because of spatial variability of soils and environments across fields (Anonymous, 2007) that are often 32 ha or greater in size. Yield monitors on combines enable the estimation of within field variability, and when a large number of field-scale replicated strip trials (56 trials) are pooled the variability in yield within fields can be similar to the variability across fields (Kyveryga et al., 2013), and for a corn-corn rotation in this study the within field variability was greater than the across field variability.

The easiest way to show the benefits of sharing data is to examine the benefits of analyses using meta-analysis and of drawing conclusions from many trials as compared with conclusions from a small number of trials. It is common for even well designed agricultural trials to have insufficient power to detect a treatment effect when there is an effect to be detected (Olkin and Shaw, 1995). Insufficient power is a common problem in research trials, with the medical field probably having the most detailed analyses of the need for large numbers of observations to make the best clinical decisions for patients and for policies (Reinhart, 2015; Button et al., 2013).

Humans are more biologically complex than agricultural fields, but agricultural fields are complex enough due to both biology and the large variety and histories of management practices across relatively small geographies to require many observations for drawing reliable conclusions (Olkin and Shaw, 1995). Many agricultural studies also frequently provide only an analysis of whether a treatment effect was present or not. The more important question for practitioners like farmers and farm advisors as well as for policy makers is the magnitude of the effect and confidence in that magnitude (Olkin and Shaw, 1995). More reliable estimates of the magnitude of an effect of a treatment in agricultural fields could be obtained by pooling data from numerous farmer networks, which is the pooling of data from many field-scale trials across many environments and years, or by a meta-analysis of small plot data. Data from field-scale trials as done by farmer networks would be preferable because the effect of spatial variability on a treatment can be estimated from trials completed on field-scale production fields but not from small-plot trials.

The advantages of meta-analysis are shown in a study using data from small-plot trials to estimate the effect of soil texture and weather on corn response to nitrogen (Tremblay, et al., 2012). Many papers have been published about the effect of soil texture and weather on corn response to nitrogen, but often the conclusions are contradictory. Contradictory results are common in research reports about the response of corn to nitrogen, and often that is because the number of locations and years where experiments were performed were insufficient to describe the variability in nitrogen response. The variability can be caused by numerous factors and soil texture and weather are two important ones cited in many studies. The meta-analysis by Tremblay et al. (2012) was able to combine 51 similar studies across 7 states in the US over 4 years with a wide range of environmental conditions and soils. These data were sufficient to quantify the effect of soil texture and weather on corn response to nitrogen, which analysis of the individual studies was unable to reliably conclude. This study, because it analyzed results from small plot trials, still leaves the reader with the questions: would the results and conclusions be different if the 51 studies had been completed on field-scale plots and not small plots, and would the magnitude of the results change if a larger number of trials were analyzed?

Combining data from field-scale trials across many farmer networks provides two large advantages over meta-analysis of trials performed on small plots. The first is field-scale trials allow the measurement of effect of spatial variability within fields on yield and a much clearer picture of how a treatment will be affected by spatial variation within fields. The second is that farmer network data would be the results from individual strips from replicated strip trials along with field history information, which is different from the means and standard deviations or mean squares for error terms typically available for meta-analysis. Results from individual strips are preferable to aggregate data because more comprehensive data analyses can be performed to more fully understand the treatment effect (Jones et al., 2009), and the data would be of much greater value to other scientists such as economists who analyze data using different techniques and hypotheses than agronomists.

An example of how a relatively small data set from one farmer network can be analyzed to quantify risk in terms of probability of a yield loss from reductions in nitrogen fertilizer applied to corn is demonstrated in Kyveryga et al. (2013). This data set contained 56 two-treatment studies over 2 years in Iowa where the nitrogen fertilizer rate was decreased by 56 kg ha^{-1} compared with the rate normally applied by the farmers who participated in the trials. Farmers often ask the question what is the risk of yield loss if they reduced their nitrogen rate by 56 kg ha^{-1}. The results showed that June rainfall had a great effect on the magnitude of a yield loss, with the year when above-normal rainfall occurred in June producing a 35% greater probability of an economic yield loss than when below-normal rainfall occurred in a corn-soybean rotation. The effect of soil organic matter differences within fields was also quantified with higher organic matter areas resulting in 20% smaller economic yield losses than areas with lower soil organic matter contents. Much more informative and richer conclusions about crop production practices across the Corn Belt could be generated from a data set that pooled data from existing farmer networks to empower analyses of hundreds or thousands of trials by agronomists and economists rather than only 56 trials. Recommendations about crop production practices, especially new practices that purport to improve the efficiency of production, from scientists could be based on reliable estimates of success of crop production practices expressed in terms of probability.

Summary

Accurate measurement of yields by yield monitors on combines provide farmers with a powerful tool to improve crop production practices. The most effective use of yield monitors is when yields are measured in replicated strip trials within a network of farmers and the results from similar trials are pooled from numerous networks. Currently it is difficult to share data among farmer networks mainly because there are no guidelines for stewardship of the data. Guidelines that provide farmers with confidence that their farm information will not become public will greatly increase the chances of creating a large data base of results from strip trials conducted by farmer networks.

The advantage of pooling data from many farmer networks is that agronomists will have sufficient data to calculate reliable probabilities describing the chances for success of new and existing crop production practices. Other scientists like economists also will have unprecedented amounts of data for analysis, which will enhance the guidelines developed by agronomists. More reliable guidelines for crop production combined with the unique local knowledge of farmers and farm advisors about field-scale production of food should improve the efficiency of crop production resulting in more profits for farmers and less pollution from unneeded inputs.

In the future agronomists will need to cooperate with statisticians to both learn how to make the best use of data sets containing results from thousands of trials over many years and to improve methods for the analysis of such data sets.

Acknowledgements

This paper was written with the support of a grant from the Environmental Defense Fund.

References

Anonymous. (2016) Data management plans. https://www.dataone.org/data-management-planning. DataOne web site. Accessed 7 June 2016.

Anonymous. (2016a) Data availability. PLOS web site. http://journals.plos.org/plosone/s/data-availability. Accessed 7 June 2016.

Anonymous. (2016b) NIH genomic data sharing policy. http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html. NIH web site. Accessed 8 June 2016.

Anonymous. (2016c) Privacy and security principles for farm data. http://www.fb.org/tmp/uploads/PrivacyAndSecurityPrinciplesForFarmData.pdf. Farm Bureau web site. Accessed 17 June 2016.

Anonymous. (2016d) Dryad publishing charges. http://datadryad.org/pages/payment. Dryad web site. Accessed 21 June 2016.

Anonymous. (2007) Interpreting small-plot and field-scale data. https://www.dekalb.ca/_uploads/documents/Agronomic_Information/CanolaArchive/west_plot_field_data.pdf. Growing Knowledge, Monsanto Technology Development Group. Accessed 18 June 2016.

Anonymous. (1996) Farmer-to-farmer networks: effective grass-roots sharing. http://www.cias.wisc.edu/farmer-to-farmer-networks-effective-grass-roots-sharing/. Accessed 17 June 2016.

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafò, M.R. (2013) Power failure: why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience.

Chapman, K, Kyveryga, P., Morris, T.F., Menke, T. (2016) Farmer network design manual: A guide for practitioners, advisors and research partners. Environmental Defense Fund. https://www.edf.org/sites/default/files/farmer-network-design-manual.pdf. Accessed 15 June 2016.

Cross, W. (2016) Libraries support data-sharing across the research lifecycle. http://lj.libraryjournal.com/2016/02/oa/libraries-support-data-sharing-across-the-research-lifecycle/. Library J. Accessed 6 June 2016.

Fecher, B., Friesike, S., Hebing, M. (2015) What drives academic data sharing? PLoS ONE 10

(2): e0118053. doi:10.1371/journal.pone.0118053.

Jahnke, L.M., Asher, A. (2014) The problem of data: data management and curation practices among university researchers. http://www.clir.org/pubs/reports/pub154/problem-of-data. Council on Library and Information Resources. Accessed 19 June 2016.

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

Kyveryga, P., Mueller, T., Paul, N., Arp, A., Reeg, P. (2015) Guide to on-farm replicated strip trials. http://www.iasoybeans.com/pdfs/OFN15_FarmerFieldGuideDigital.pdf. Accessed 18 June 2016.

Metcalf, J. (2015) Data management plan: A background report. http://bdes.datasociety.net/wp-content/uploads/2015/05/DMPReport.pdf. Council for Big Data, Ethics, and Society. Accessed 7 June 2016.

Olkin, I., Shaw, D. V. (1995) Meta-analysis and its applications in horticultural science. HortSci. 30:1343-1348.

Puniewska, M. (2014) Scientists have a data sharing problem.

http://www.theatlantic.com/health/archive/2014/12/scientists-have-a-sharing-problem/383061/. The Atlantic. Accessed 7 June 2016.

Savage, C.J., Vickers, A.J. (2009) Empirical study of data sharing by authors publishing in PLoS journals. PloSOne 4: e7078. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0007078. Accessed 15 June 2016.

Reinhart, A. (2015) Statistics done wrong. The woefully complete guide. San Francisco, CA. No Starch Press. ISBN: 978-1-59327-620-1.

Roche DG, Lanfear R, Binning SA, Haff TM, Schwanz LE, et al. (2014) Troubleshooting Public Data Archiving: Suggestions to Increase Participation. PLoS Biol 12(1): e1001779. doi: 10.1371/journal.pbio.1001779.

Tenopir, C., Dalton, E.D., Allard, S. Frame, M., Pjesivac, I., Birch, B., Pollock, D., Dorsett, K. (2015) Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. PLosOne. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0134826. Accessed 16 June 2016.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, R., Manoff, M., Frame, M. (2011) Data sharing by scientists: practices and perceptions. PLoSOne 6: e21101. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101. Accessed 16 June 2016.

Tremblay, N., Bouroubi, Y.M., Bélec, C., Mullen, R.W., Kitchen, N.R., Thomason, W.E., Ebelhar, S., Mengel, D.B., Raun, W.R., Francis, D.D., Vories, E.D., and Ortiz-Monasterio, I. (2012) Corn response to nitrogen is influenced by soil texture and weather. Agron J. 104: 1658-1671. https://dl.sciencesocieties.org/publications/aj/pdfs/104/6/1658?search-result=1. Accessed 21 June 2016.