



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*



Big Data in Farming: Why Matters!

Steve Sonka and Yu-Tien Cheng

University of Illinois

November 12, 2015

farmdoc daily (5):211

Recommended citation format: Sonka, S., and Y.-T., Cheng. "Big Data in Farming: Why Matters!" *farmdoc daily* (5):211, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, November 12, 2015.

Permalink: <http://farmdocdaily.illinois.edu/2015/11/big-data-in-farming-why-matters.html>

Big Data applications generally create predictions based on analysis of what has occurred. Uncertainty in farming, based in biology and weather, means that what has happened often is not a good predictor of what will happen. This means that if the science of agriculture (the why) is not integrated within the sector's Big Data applications, serious mistakes could result. Figure 1 illustrates the concept of integrating what and why. The concept will be further explained later. This is the third of a six-part series on [Big Data and Agriculture](#).

Figure 1. Fusion of Data Concepts!

fdd



Photo credit: Stephen Mcsweeny/Shutterstock.com

We request all readers, electronic media and others follow our citation guidelines when re-posting articles from farmdoc daily. Guidelines are available [here](#). The farmdoc daily website falls under University of Illinois copyright and intellectual property rights. For a detailed statement, please see the University of Illinois Copyright Information and Policies [here](#).

Do we really need to care about why?

In the first article of this series, we cited an IBM forecast that the data which exists in the world today will be doubled in two years -- on November 12, 2017 ([IBM 2015](#)). With the promise of all this data, it sometimes seems that soon we're going to know what happened everywhere. And all that data will "surely" enable us to come up with greatly improved recommendations and decisions.

The title of a recent book, *BIG DATA: A Revolution That Will Transform How We Live, Work, and Think*, illustrates the belief that this thing called Big Data is a really BIG deal (Mayer-Schönberger and Cukier 2013). Consider the title — transforming how we live, work, and think — what else is there? In that book, the authors predict:

Society will need to shed some of its obsession for causality in exchange for simple correlations: not knowing why but only what. This overturns centuries of established practices and challenges our most basic understanding of how to make decisions and comprehend reality (Mayer-Schönberger and Cukier 2013, 7).

This is a bold statement. The kind people like to make when they're trying to sell their own books.

But is that really the case? If we just have lots of data about something, will that always (or maybe even often) help us to recommend future actions? Let's digress for a minute to consider relationships between nutrition and length of life. Figures 2 and 3 illustrate (in an admittedly light-hearted way) why just having data about complex issues may not lead to the best conclusion. First, read Figure 2 to review a number of relationships linking diet and length of life. Then, Figure 3 will provide the obvious recommendation based on those relationships.

Figure 2. Diet and Length of Life Around the World

fdd

- The Japanese eat very little fat
 - *suffer fewer heart attacks than do the British or Americans*
- The French eat a lot of fat
 - *suffer fewer heart attacks than do the British or Americans*
- The Italians drink excessive amounts of red wine
 - *suffer fewer heart attacks than do the British or Americans*
- The Japanese drink very little red wine
 - *suffer fewer heart attacks than do the British or Americans*
- The Germans drink a lot of beer and eat lots of sausages
 - *suffer fewer heart attacks than do the British or Americans*

Figure 3. So...The Recommendation is:

fdd

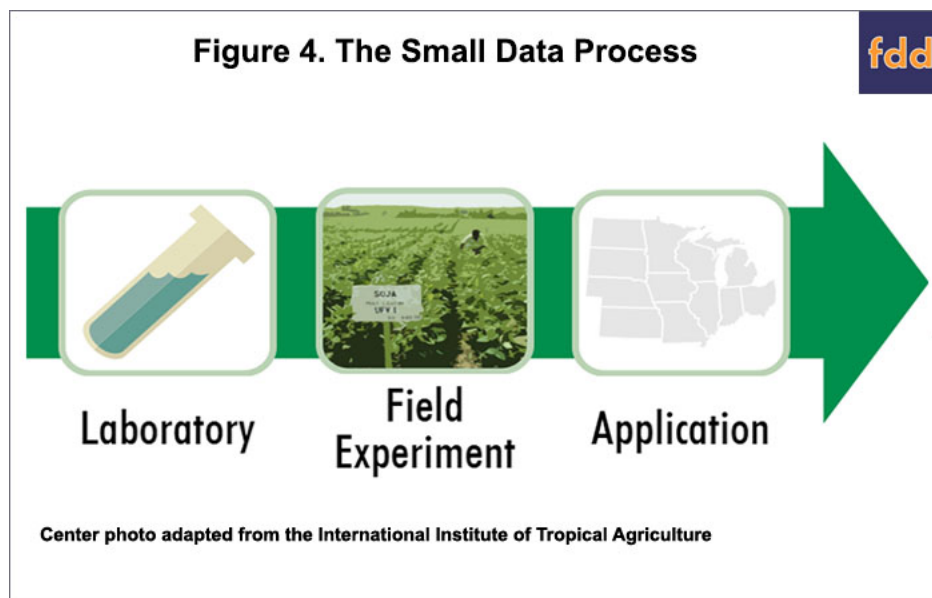
EAT what you want,
Apparently it's
SPEAKING ENGLISH
that kills you!

Of course, any good Big Data enthusiast would quickly assert that the problem with this recommendation is that we didn't consider a sufficient number of variables. We were lacking in the Variety dimension of Big Data. True; but the larger point is that we can easily mislead ourselves when we're faced only with lots of data linked to decisions regarding complex, uncertain systems.

Production agriculture is complex, where biology, weather, and human actions interact. Indeed agriculture has its own example of the dangers of inferring causality based upon observations of what happened. In the 1870s, farmers in both the United States and Australia opened new farm areas where rainfall was thought to be too marginal for sustained crop production. In both instances, however, increases in rainfall occurred as the area of crop production increased. The resulting conclusion was the famous assertion that, "rainfall follows the plow". That relationship was thought to explain how the Great American Desert had become fertile ([Wikipedia 2014](#)). Commercial interests extensively promoted that belief and settlers followed. Unfortunately, when normal, dry conditions returned, considerable human and environmental hardship followed.

"Small Data"—the key to modern agriculture

Fortunately, we don't have to rely **only** upon observations of what has happened. Modern agriculture is predicated on knowing and applying the why that drives crop and livestock production. Indeed, it is this "small data" process, illustrated in Figure 4, which fostered the advances in agriculture over the last 100 years.



The process starts with lab research employing the scientific method as a systematic process to gain knowledge through experimentation. Indeed the scientific method is designed to ensure that the results of an experimental study did not occur just by chance (Herren 2014). However, results left in the lab don't lead to innovation and progress in the farm field. In the United States, the USDA, Land Grant universities, and the private sector have collaborated to exploit scientific advances. A highly effective, but distributed, system emerged where knowledge gained in the laboratory was tested and refined on experimental plots and then extended to agricultural producers.

Fusing of Small and Big Data

Sometimes, knowing only what likely is sufficient. At 8:30 am on any weekday morning, traffic will be jammed on the Kennedy Expressway entering downtown Chicago. Information that tells us when the traffic starts to open up is useful.

Knowing, at increasing levels of precision, what happened in the field or in animal facilities also does have value. However, also knowing why is critically important in agriculture. On the one hand, what vs. why is a challenge. But, it also presents a major opportunity.

Taking advantage of the synergies available from fusing the knowledge residing in both why and what seems highly attractive. This approach is particularly intriguing in the context of the dynamic world of agriculture. Weather events vary from year to year and pests evolve in their location and behavior. Advances in genetics and research continue to provide enhanced but differing capabilities. Consumer preferences, market demands, and societal expectations can be volatile in their behavior.

Historically the research necessary to determine the why's of crop and livestock production were conducted in the public sector and communicated through public extension services. Over the latter part of the last century, the private sector's role expanded to include provision of those functions.

The organizational challenges and opportunities associated with achieving effective collaboration among farms, agribusiness and technology firms, and the public sector are significant. We will focus on this topic in the final article of this series.

References

Herren, R.V. *The Science Of Agriculture A Biological Approach 4th Ed.* Independence: Cengage Learning, 2014.

IBM. *What is Big Data?* Last modified September 29, 2015 and accessed November 9, 2015. <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>.

Mayer-Schönberger, V. and K. Cukier. *BIG DATA: A Revolution That Will Transform How We Live, Work, and Think.* Boston: Eamon Dolan/Houghton Mifflin Harcourt, 2013.

Wikipedia. *Rain Follows the Plow.* Last modified September 26, 2014, and accessed November 9, 2015. https://en.wikipedia.org/wiki/Rain_follows_the_plow.