



Big Data: More Than a Lot of Numbers!

Steve Sonka and Yu-Tien Cheng

University of Illinois

October 29, 2015

farmdoc daily (5):201

Recommended citation format: Sonka, S., and Y.-T., Cheng. "Big Data: More Than a Lot of Numbers!" *farmdoc daily* (5):201, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, October 29, 2015.

Permalink: <http://farmdocdaily.illinois.edu/2015/10/big-data-more-than-a-lot-of-numbers.html>

Big Data and Agriculture

Big Data — the current buzzword of choice. Today it's very easy to be overwhelmed by the hype promoting Big Data. Farm media, newspapers and general media, and conference speakers all extol the future transforming effects of Big Data, stressing that "**Big Data** will be essential to our future, whatever **it is**." The goal of this article, and the series of five that follow, is to begin to unravel that "whatever **it is**" factor for agriculture.

We'll definitely explore "whatever it is" from a managerial, not a computer science, perspective. Potential implications for agriculture will be the primary emphasis of the following set of articles:

1. *Big Data: More Than a Lot of Numbers!* This article emphasizes the role of analytics enabling the integration of various data types to generate insights. It stresses that the "Big" part of Big Data is necessary but it's the "Data" part of Big Data that's likely to affect management decisions.
2. *Precision Ag: Not the Same as Big Data But...* Today, it's easy to be confused by the two concepts, Precision Ag and Big Data. In addition to briefly reviewing the impact of Precision Ag, this article stresses that Big Data is much more than Precision Ag. However, Precision Ag operations often will generate key elements of the data needed for Big Data applications.
3. *Big Data in Farming: Why Matters!* Big Data applications generally create predictions based on analysis of what has occurred. Uncertainty in farming, based in biology and weather, means that the science of agriculture (the Why) will need to be integrated within many of the sector's Big Data applications.
4. *Big Data: Alive and Growing in the Food Sector!* Big Data already is being extensively employed at the genetics and consumer ends of the food and ag supply chain. This article will stress the potential for capabilities and knowledge generated at these levels to affect new opportunities within production agriculture.

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](#).

5. *A Big Data Revolution: What Would Drive It?* Management within farming historically has been constrained by the fundamental reality that the cost of real-time measurement of farming operations exceeded the benefits from doing so. Sensing capabilities (from satellites, to drones, to small-scale weather monitors, to soil moisture and drainage metering) now being implemented will materially lessen that constraint. Doing so will create data streams (or is it floods?) by which Big Data applications can profoundly alter management on the farm.
6. *A Big Data Revolution: Who Would Drive It?* Over the last 30 years, novel applications of information technology have caused strategic change in many sectors of the economy. This article draws on those experiences to inform our thinking about the potential role of Big Data as a force for change in agriculture.

Big Data: More Than a Lot of Numbers!

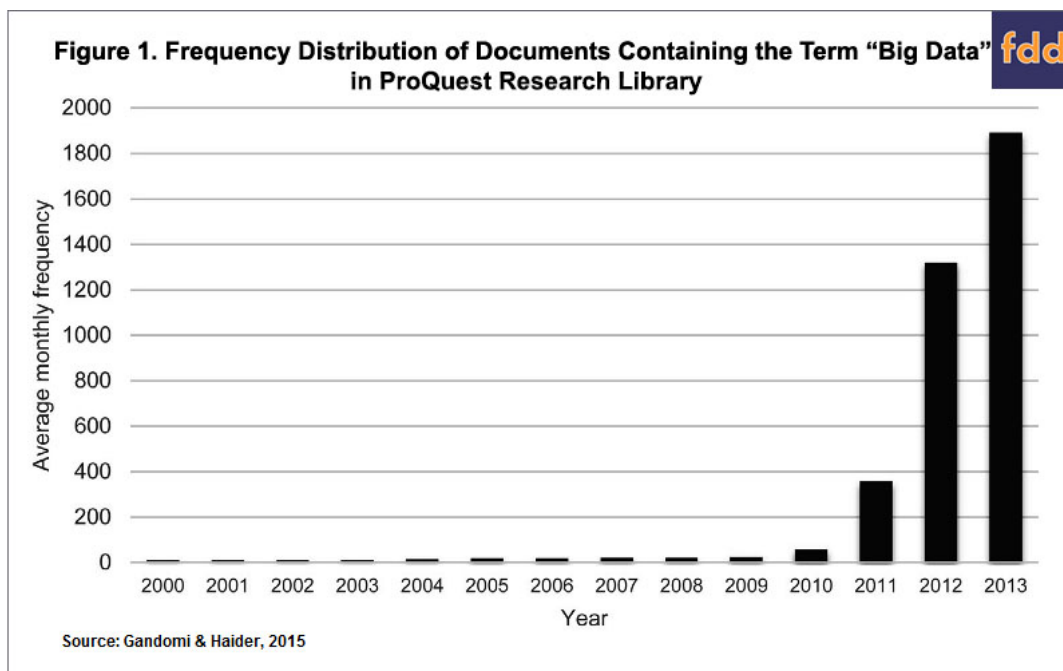
Innovation has been critical to increased agricultural productivity and to support of an ever increasing global population. To be effective, however, each innovation had to be understood, adopted, and adapted by farmers and other managers. Although Big Data is relatively new, it is the focus of intense media speculation today. However, it is important to remember that Big Data won't have much impact unless it too is understood, adopted and adapted by farmers and other managers. This article provides several perspectives to support that process.

Big Data Defined

"90% of the data in the world today has been created in the last two years alone" (IBM, 2012).

In recent years, statements similar to IBM's observation and associated predictions of a Big Data revolution have become increasingly more common. Some days it seems like we can't escape them!

Actually, Big Data and its hype are relatively new. As shown in Figure 1, use of the term, Big Data, was barely noticeable prior to 2011. However, the term's usage literally exploded in 2012 and 2013, expanding by a factor of 5 in just two years.



With all new concepts, it's nice to have a definition. Big Data has had more than its fair share. Two that we find helpful are:

- The phrase "big data" refers to large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future ([National Science Foundation, 2012](#)).
- Big Data is high-volume, -velocity, and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making ([Gartner IT Glossary, 2012](#)).

These definitions are impressive. However, they really don't tell us how Big Data will empower decision makers to create new economic and social value.

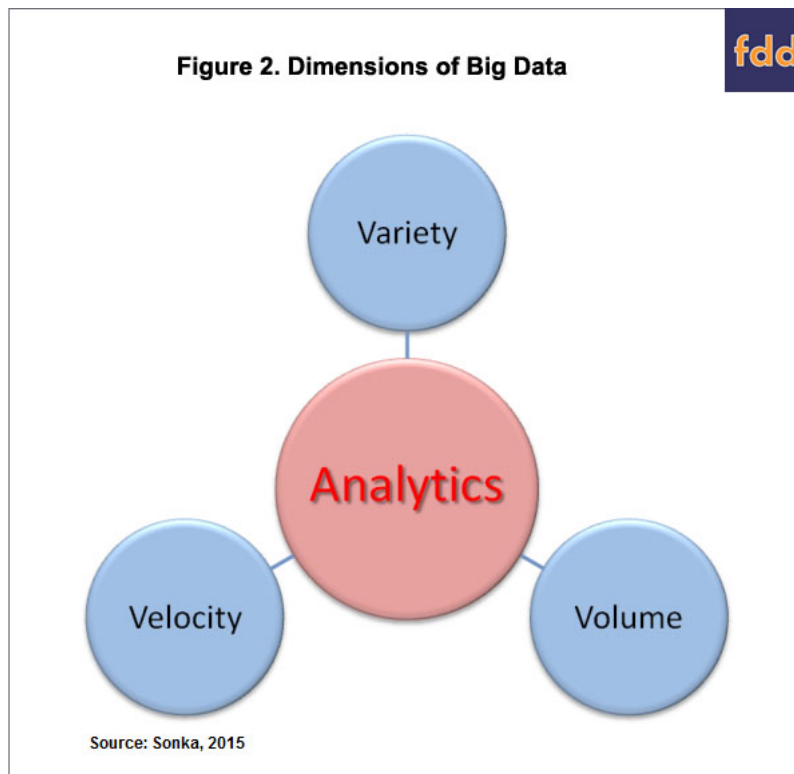
From Technology to Value

In the next few paragraphs, we'll move beyond those definitions to explore how application of Big Data fosters economic growth. In this article, we'll present non-ag examples because today there is more experience outside of agriculture. The following articles in this series will focus on agriculture.

Big Data generally is referred to as a singular thing. It's not! In reality, Big Data is a capability. It is the capability to extract information and craft insights where previously it was not possible to do so. Advances across several technologies are fueling the growing Big Data capability. These include, but are not limited to computation, data storage, communications, and sensing.

These individual technologies are "cool" and exciting. However, sometimes a focus on cool technologies can distract us from what is managerially important.

A commonly used lens when examining Big Data is to focus on its dimensions. Three dimensions (Figure 2) often are employed to describe Big Data: **Volume**, **Velocity**, and **Variety**. These three dimensions focus on the nature of data. However, just having data isn't sufficient. **Analytics** is the hidden, "secret sauce" of Big Data. Analytics refers to the increasingly sophisticated means by which analysts can create useful insights from available data.



Now let's consider each dimension individually:

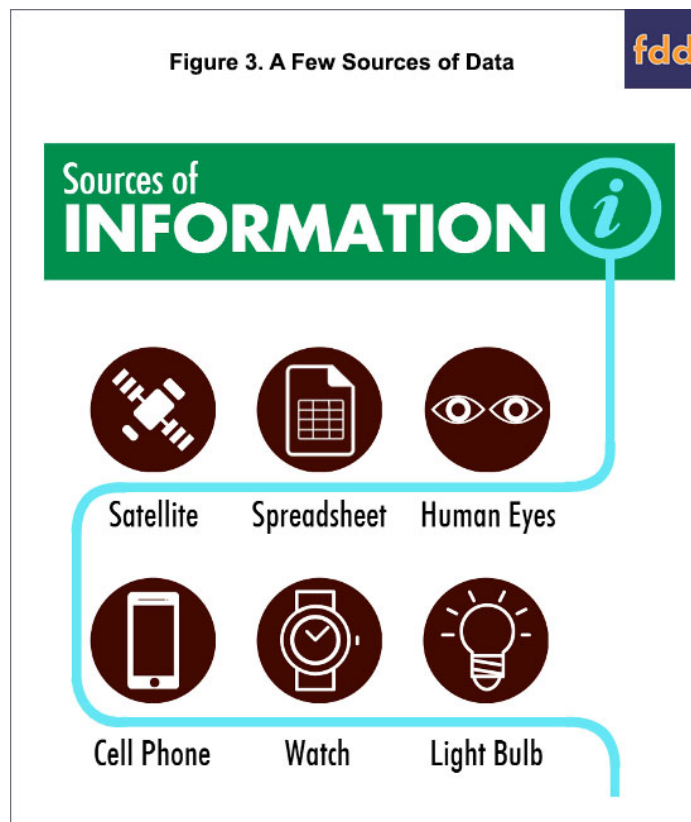
Interestingly, the **Volume** dimension of Big Data is not specifically defined. No single standard value specifies how big a dataset needs to be for it to be considered "Big". It's not like Starbucks; where the Tall cup is 12 ounces and the Grande is 16 ounces. Rather, Big Data refers to datasets whose size exceeds the ability of the typical software used to capture, store, manage, and analyze.

This perspective is intentionally subjective and what is "Big" varies between industries and applications. An example of one firm's use of Big Data is provided by GE — which now collects 50 million pieces of data from 10 million sensors everyday (Hardy, 2014). GE installs sensors on turbines to collect information on the "health" of the blades. Typically, one gas turbine can generate 500 gigabytes of data daily. If use of that data can improve energy efficiency by 1%, GE can help customers save a total of \$300 billion (Marr, 2014)! The numbers and their economic impact do get "Big" very quickly.

The **Velocity** dimension refers to the capability to acquire, understand, and respond to events as *they occur*. Sometimes it's not enough just to know what's happened; rather we want to know what's happening. We've all become familiar with real-time traffic information available at our fingertips. Google Map provides live traffic information by analyzing the speed of phones using the Google Map app on the road (Barth, 2009). Based on the changing traffic status and extensive analysis of factors that affect congestion, Google Map can suggest alternative routes in real-time to ensure a faster and smoother drive.

Variety, as a Big Data dimension, may be the most novel and intriguing. For many of us, our image of data is a spreadsheet filled with numbers meaningfully arranged in rows and columns.

With Big Data, the reality of "what is data" has wildly expanded. The lower row of Figure 3 shows some newer kinds of sensors in the world, from cell phones, to smart watches, and to smart lights. Cell phones and watches can now monitor users' health. Even light bulbs can be used to observe movements, which help some retailers to detect consumer behaviors in stores to personalize promotions (Reed, 2015). We even include human eyes in Figure 3, as it would be possible to track your eyes as you read this article.



The power of integrating across diverse types and sources of data is commercially substantial. For example, UPS vehicles are installed with sensors to track the engine performance, car speed, braking, direction, and more (van Rijmenam, 2014). By analyzing these and other data, UPS is able to not only monitor the car engine and driving behavior but also suggest better routes, leading to substantial savings of fuel (Schlangenstein, 2013).

So, Volume, Variety, and Velocity can give us access to lots of data, generated from diverse sources with minimal lag times. At first glance that sounds attractive. Fairly quickly, however, managers start to wonder, what do I **do** with all this stuff? Just acquiring more data isn't very exciting and won't improve agriculture. Instead, we need tools that can enable managers to improve decision-making; this is the domain of **Analytics**.

One tool providing such capabilities was recently unveiled by the giant retailer, Amazon (Bensinger, 2014). This patented tool will enable Amazon managers to undertake what it calls "anticipatory shipping", a method to start delivering packages even before customers click "buy". Amazon intends to box and ship products it expects customers in a specific area will want but haven't yet ordered. In deciding what to ship, Amazon's analytical process considers previous orders, product searches, wish lists, shopping-cart contents, returns, and even how long an Internet user's cursor hovers over an item.

Analytics and its related, more recent term, data science, are key factors by which Big Data capabilities actually can contribute to improved performance, not just in retailing, but also in agriculture. Such tools are currently being developed for the sector, although these efforts typically are at early stages.

So What?

In this discussion, we explored the dimensions of Big Data — 3Vs and an A. The Volume dimension links directly to the "Big" component of Big Data. Variety, Velocity and Analytics relate to the "Data" aspect. While Volume is important, strategic change and managerial challenges will be driven by Variety, Velocity, and especially Analytics. Unfortunately, media and advertising tend to emphasize Volume; it's easy to impress with really, really large numbers. But farmers and agricultural managers shouldn't be distracted by statistics on Volume.

Big Data's potential doesn't rest on having lots of numbers or even having the world's largest spreadsheet. Instead, the ability to integrate across numerous and novel data sources is key. The point of doing this is to create new managerial insights that enable better decisions. While Volume and Variety are necessary, Analytics is what allows for fusion across data sources and new knowledge to be created.

Emphasizing the critical role of Variety of data sources and Analytics capabilities is particularly important for production agriculture. Individual farms and other agricultural firms aren't likely to possess the entire range of data sources needed to optimize value creation. Further, sophisticated and specialized Analytics competencies will be required. To be effective, however, the computer science competencies also need to be combined with knowledge of the business and science aspects of agricultural production.

At times this sounds complicated and maybe threatening. Visiting with a farmer from Ohio about this topic recently, he made a comment that is helpful in unraveling this complexity. He noted that effective use of Big Data for him as a Midwestern farmer is mainly about relationships. The relevant question is, "Which input and information suppliers and customers can provide the Big Data capabilities for him to optimize his decisions?" And he noted, "For farmers, managing those relationships isn't new!"

References

Barth, D. "The bright side of sitting in traffic: Crowdsourcing road congestion data." Posted August 25, 2009, and accessed October 27, 2015.
<https://googleblog.blogspot.com/2009/08/bright-side-of-sitting-in-traffic.html>

Bensinger, G. "Amazon wants to ship your package before you buy it." *The Wall Street Journal*. Posted January 17, 2014, and accessed October 27, 2015.
<http://blogs.wsj.com/digits/2014/01/17/amazon-wants-to-ship-your-package-before-you-buy-it/>

Gandomi, A. & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35, 137-144.

Gartner IT Glossary. (2012). *Big Data*. Retrieved from <http://www.gartner.com/it-glossary/big-data>

Hardy, Q. "G.E. opens its Big Data platform." *The New York Times*. Posted October 9, 2014, and accessed October 27, 2015. http://bits.blogs.nytimes.com/2014/10/09/ge-opens-its-big-data-platform/?_r=2

IBM. (2012). *What is big data?* Retrieved from
<https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>

Marr, B. "How GE is using Big Data to drive business performance." Posted September 3, 2014, and accessed October 27, 2015.
<http://www.smartdatacollective.com/bernardmarr/229151/how-ge-using-big-data-drive-business-performance>

National Science Foundation. (2012). *Core techniques and technologies for advancing big data science & engineering*. Retrieved from <http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.htm>

Reed, J. "How Philips Lighting mastered smart lights – and turned a \$60 light bulb into a winner." Posted March 18, 2015, and accessed October 27, 2015.
<http://diginomica.com/2015/03/18/how-philips-lighting-mastered-smart-lights-and-turned-a-60-light-bulb-in-to-a-winner/#.Vi3FedKrTIX>

Schlangenstein, M. "UPS crushes data to make routes more efficient, save gas." *Bloomberg*. Posted October 30, 2013, and accessed October 27, 2015.
<http://www.bloomberg.com/news/articles/2013-10-30/ups-uses-big-data-to-make-routes-more-efficient-save-gas>

Sonka, S. (2015). Big data: From hype to agricultural tool. *Farm Policy Journal*, 12(1), 1-9.

van Rijmenam, M. "Why UPS spends over \$ 1 Billion on Big Data annually." Posted May 23, 2014, and accessed October 27, 2015. Retrieved from
<https://datafloq.com/read/ups-spends-1-billion-big-data-annually/273>