International Food and Agribusiness Management Review
Volume 17 Issue 1, 2014

Big Data and the Ag Sector: More than Lots of Numbers

Steve Sonka

Director, ADM Institute for the Prevention of Postharvest Loss, Professor Emeritus of Agricultural Strategy, University of Illinois, 807 S. Wright Street, Suite 320, Champaign, Illinois, 61820, USA

Abstract

It seems that one can't go through a work day without seeing some mention of Big Data, its application and its potential to have unprecedented impact. The potential for Big Data application in the agricultural sector is examined. The role of analytics and the variety and velocity characteristics of Big Data as they can apply to the sector are stressed. Integration of data and analysis across business and government entities will be needed for successful implementation. The eventual impact of Big Data within the agricultural sector likely will require both organizational and technological innovation.

Keywords: Big Data, agricultural supply chain integration, information and communication technology

Corresponding author: Tel: + 1 217. 333. 5115
Email: S. Sonka: ssonka@illinois.edu

Over the last three decades, the application of information and communication technologies (ICT) has had marked impact across society and the economy. Changes fueled by ICT adoption are apparent to us today and the processes by which those changes occurred are a tacit part of our experience base. As we moved through that adoption process, however, the extent of adoption and its eventual effects were not nearly so clear. We asked questions such as:

In the 1980s,

- What is a microcomputer?
- If my office has a Selectric typewriter, why would I need to do word processing?
- Isn't 32K RAM (Random Access Memory) plenty?

In the 1990s,

- I have voice mail for my phone, why would I use e-mail?
- What is this Internet thing?
- Would farmers actually pay for GPS-based yield monitors?

In the 2000s,

- Why would I buy a book on-line when the bookstore is just down the road?
- I get calls on my current cell phone, why buy something called a Smartphone?
- Google it?

Today, ICT-based advances continue to offer opportunities and challenges. One of the most talked about, for business, government, and society, is called "***Big Data***". (A term often followed by either the phrase –"whatever that is" or by a measure expressed as "****bytes", which then is explained as equivalent to some multiple of the information stored in the Library of Congress.)

Big Data is perceived to be as relevant for agriculture as it is for the rest of the economy. Indeed, Padmasree Warrior, Chief Technology and Strategy Officer for Cisco Systems (Kirkland 2013), notes:

> *In the next three to five years, as users we'll actually lean forward to use technology more versus what we had done in the past, where technology was coming to us. That will change everything, right? It will change health care;* ***it could even change farming****. There are new companies thinking about how you can farm differently using technology; sensors connected that use water more efficiently, use light, sunlight, more efficiently.*

The purpose of this article is to examine the Big Data phenomenon and to explore its implications for agriculture. From a managerial perspective it is important to stress that Big Data encompasses much more than just lots of numbers. It provides the potential for managers to have access to explicit information and decision making capabilities that have not been available previously. For some, the resulting innovations likely offer opportunity but, for others, they represent threats.

The Big Data phenomenon is driven by stunning and exciting advances in technology. Taking advantage of these advances in the ag sector could require new organizational linkages to be formed – between suppliers and customers and among competitors. The effective evolution of such linkages will materially affect the manner and extent of Big Data's effect on the sector.

The remainder of this report is comprised of the following three sections:

- Exploring Big Data and characteristics of its business application to date
- Big Data and agriculture: vision and/or hallucination?
- Wrapping it up

Exploring Big Data and Characteristics of its Business Application to Date

Although of great potential importance, the Big Data phrase is only the latest buzz word to capture media attention. Interestingly, there seems to be a continual pattern in society's response to such phenomenon; we tend to overestimate the initial impact and underestimate the long run effect. Both perceptions can lead to difficulties for agricultural managers as they fashion the most effective response to innovation possibilities.

The Big Data concept seems to be clearly in the initial overemphasis stage, where media hype and advertising are capturing significant attention. In this section, we'll attempt to present available information without being overwhelmed by the enthusiasm of the moment. The examples noted in this section will not be limited to applications in production agriculture or the food chain. Linking the concepts noted here to potential application in agriculture will be the focus of the paper's following section.

In this section, we'll focus on the following three components of Big Data:

- What is Big Data and why is it different?
- Example applications
- What are the managerial lessons so far?

What is Big Data and Why is it Different?

The reference source of choice for today's college student (Wikipedia 2013) highlights six examples of Big Data use:

- Amazon.com handles millions of back-end operations every day, as well as queries from more than half a million third-party sellers.
- Walmart handles more than 1 million customer transactions every hour, which is imported into databases estimated to contain more than 2.5 petabytes (2560 terabytes) of data – the equivalent of 167 times the information contained in all the books in the US Library of Congress.
- Facebook handles 50 billion photos from its user base.

- FICO Falcon Credit Card Fraud Detection System protects 2.1 billion active accounts world-wide.
- The volume of business data worldwide, across all companies, doubles every 1.2 years, according to estimates.
- Windermere Real Estate uses anonymous GPS signals from nearly 100 million drivers to help new home buyers determine their typical drive times to and from work throughout various times of the day.

These examples illustrate, but don't completely capture, the diversity of Big Data applications. Looking behind such examples, three dimensions are typically used to describe the Big Data phenomenon; **Volume**, **Velocity**, and **Variety**. Each dimension present challenges for data management and for exploiting opportunities to advance business decision making. These three dimensions focus on the nature of data, however, as we'll see just having data isn't sufficient. Therefore an overview of **Analytics** also will be presented in this section.

We'll employ the concepts of a recent paper of the McKinsey Global Institute to describe the dimension of Big Data (Manyika et al. 2011). Interestingly, the **volume** dimension of Big Data is not delineated in quantitative terms. Rather, Big Data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze. This definition is intentionally subjective and incorporates a moving definition of how big a dataset needs to be in order to be considered big. As technology advances over time, the size of datasets that qualify as Big Data will also increase. Also the definition can vary by sector, depending on what kinds of software tools are commonly available and what sizes of datasets are common in a particular industry. With those caveats, Big Data in many sectors today will range from a few dozen terabytes to multiple petabytes (thousands of terabytes).

The **velocity** dimension refers to the capability to acquire, understand, and interpret events **AS** they occur. For analysts interested in retailing, anticipating the level of sales is critically important. McAfee and Brynjolfsson (2012) report on an effort to monitor mobile phone traffic to infer how many people were in the parking lots of a key retailer on Black Friday – the start of the holiday shopping season -- as a means to estimate retail sales.

Variety, as a dimension of Big Data, may be the most novel and intriguing of these three characteristics. For many senior managers, the personal computer freed us from the tyranny of the IT department's chokehold on data. In our experience, however, data refers to numbers meaningfully arranged in rows and columns that can then be summarized in appropriate standardized reports. Usually these numbers summarized operating and financial performance.

For Big Data, the concept of "what is data" is wildly expanded. We now refer to the "Internet of Things"—acquisition of data from activities in the physical world, where sensors embedded in physical objects continually report on their status. In the Internet of Things, processes continually monitor and report their own activities and products are themselves source of information – of what they are, where they are and where they're headed. The potential role of nanotechnology-based sensors will be critical to low-cost, widespread availability of sensor capabilities (Lu and Bowles 2013).

Of equal importance is the capability for Big Data efforts to extract value from analysis of unstructured, qualitative data. Searching text files is not new. Previously, however, such efforts had to be carefully structured, with precise search terms and limitations on the amount of text examined. Big Data tools enable analysts to explore massive quantities of text and allow for the analysis itself to identify the relevant descriptors within the information.

ICT advances have materially affected what is data and the ability of businesses to capture, transmit and store data. But, we still have to make sense of the information inherent in these massive amounts of diverse data to learn how to make better decisions and implement new opportunities. **Analytics** (often with information, business, or social as an adjective) is the term now used to describe this process.

A comparison of shopping for a book in a physical store versus shopping on-line illustrates the change of perspective. Tracking, of which books sold and which didn't, could be done effectively in a physical bookstore. Combined with a loyalty program, that information could be linked to characteristics of the individual purchaser. This information needs to be captured by the on-line bookseller as well. However, the on-line bookseller can <u>track and influence</u> the shopping experience of each customer. The on-line bookseller can track what books were examined but not purchased and how the customer navigated through the site. The customer, while shopping, can be offered promotions and suggestions – linked to that customer's characteristics. Further the effectiveness of the suggestions and promotions can be assessed. Learning from and influencing the shopping experience is driven by development and implementation of algorithms in the process referred to as information analytics.

Of course, brick and mortar retailers have the potential to employ ICT applications in novel ways to develop improved insight regarding their in-person customers (Clifford and Hardy 2013). Tracking cell phone signals as customers move through the store can give an indication of what is of most interest. Facial recognition software may allow clerks to anticipate the shopper's mood and thereby improve their interaction with the customer. Interestingly some in-person customers, when informed that in-store tracking is occurring, respond negatively to the practice even though they report little concern with website tracking.

Example Applications

We just explored the differing perspectives available to a physical bookstore versus its on-line competitor, where the insights gained regarding each customer were stacked in favor of the on-line offering. In Exhibit 1, examples of manufacturing applications are described (Hagerty 2013).

Exhibit 1. Monitoring Manufacturing Processes.

> Companies' pursuit of "big data"—collecting and crunching ever larger amounts of information—is often thought of as another way to figure out exactly what customers want. But big data is also a means of measuring millions of little things in factories, such as how many times each screw is turned.
>
> That is what Raytheon Co. is doing at a new missile plant in Huntsville, Ala. If a screw is supposed to be turned 13 times after it is inserted but is instead turned only 12 times, an error message flashes and production of the missile or component halts, says Randy Stevenson, a missile-systems executive at Raytheon. Improvising with a defective screw or the wrong size screw isn't an option, he says. "It's either right or it's not right."
>
> At Harley-Davidson Inc.'s newly renovated motorcycle plant in York, Pa., software keeps a constant record of the tiniest details of production, such as the speed of fans in the painting booth. When the software detects that fan speed, temperature, humidity or some other variable is drifting away from the prescribed setting, it automatically adjusts the machinery.
>
> "It allows us to be more consistent," says John Dansby II, vice president for global manufacturing at the motorcycle maker. In the past, he says, operators had a bit of leeway on paint jobs and each could do the work in a slightly different way. Now it is supposed to be an exact science, not an art.
>
> Harley has also used the software to find bottlenecks that could keep it from its goal of completing a motorcycle every 86 seconds. Recently, by studying the data, Harley managers determined that installation of the rear fender was taking too long. They changed a factory configuration so those fenders would flow directly to the assembly line rather than having to be put on carts and moved across an aisle.

Recently, public attention was fixated on another Big Data application, one that didn't directly involve manufacturing or consumers. Instead it was revealed that the National Security Agency was efficiently able to parse millions of phone, text and online conversations for information. Hickens (2013) describes the advances in technology which facilitated this Big Data application. In particular, software and hardware advances are noted which drive rapid assessment of massive amounts of unstructured data. These advances often employed tools developed for commercial application.

Managerial Lessons

The reality that ICT-based advances can enhance business operations and disrupt industries is not a recent event. Indeed, for managers in the 1990s, the challenge of dealing with and taking advantage of the "knowledge economy" parallels the attention devoted to Big Data today. At that time, there was considerable speculation regarding the advent and potential effects of the knowledge economy. However, some key concepts were identified that have proven to have long-run applicability. Central among these findings is the notion that, while the new technology is essential, the capability to exploit the learning that results from the use of the new technology is the important element of the process.

Extensive use of information technology has redefined industries throughout the economy. Although the effects often have been described in some detail, the underlying mechanisms that fuel industry transformation have been less well understood. Sampler (1997) provides an important analysis of these underlying mechanisms. His analysis stresses that, although industry transformation may be the result, we need to understand the impact of information technology at the level of the individual transaction. Two key transaction characteristics are identified:

- Separability refers to the extent to which specific information attributes can be captured in association with each transaction and

- Aggregation potential refers to the extent to which those information attributes can be leveraged to gain economic value beyond the purpose of the original transaction.

Traditionally an economic transaction is perceived as the exchange of a good or service for cash. The information attributes that *must* be captured in that exchange are relatively minimal; the amount of cash and the quantity of the product. Clearly the introduction of low cost information systems has materially altered the nature and amount of attributes that are routinely captured in many of today's economic exchanges. Considerable effort is now expended to identify the purchaser and to profile the purchaser demographically.

But information technologies are now employed in settings that require us to alter our perception of what constitutes an economically relevant transaction. Real time sensors, for example, can monitor engine tire wear so that each revolution of the tire is a transaction. When equipped with communication capabilities, such monitors can alert decision makers of the potential for critical problems before these problems occur.

The second transaction characteristic is aggregation potential. Knowing the purchase habits of one consumer is interesting but that information provides little economic value. However, being able to accumulate and analyze the purchase behaviors of many consumers can have considerable value. In the case of a sensor tracking engine performance, there is value (which must be at least greater than the cost of employing the sensor system) in knowing the status of that one engine. Again, however, there are additional benefits available if that data can be accumulated, analyzed, and used to predict and enhance future performance. These examples illustrate the key role of the aggregation potential characteristic as the use of information technology redefines industries.

Aggregation potential typically requires sophisticated analysis, extensive communications and the ability to capture returns from wide-spread application of the algorithms defined. As detailed by Shapiro and Varian (1999), these characteristics result in significant economies of scale on the "supply-side" of information economics. First-mover advantages, therefore, accrue to the first firm to effectively create a system that can exploit aggregation potential. These "winner-takes-all" effects have interesting implications for future industry structure where they occur. These scale economics (of information aggregation) naturally transfer to the realm of Big Data.

A recent analysis conducted by the Economist Intelligence Unit indicates that business leaders feel there is a strong link between use of Big Data and economic performance (Economist Intelligence Unit). Based upon surveys with several hundred executives from across the globe and industry sectors, it was found that top-performing companies:

- Process data more rapidly than their peers,
- Acquire more data from outside their internal operations, and
- Use the data in more functions across the firm.

The overall lesson is that more successful firms exploit Big Data by focusing on business priorities.

Barton and Court (2012) worked with dozens of firms in data rich industries. They identify three supportive capabilities as essential to achieving success in the application of Big Data. These are the ability to:

- Identify, combine, and manage multiple sources of data,
- Create advanced analytic models for predicting and optimizing outcomes, and
- Transform the organization so that data and models actually yield better decisions.

Big Data and Agriculture: Vision and/or Hallucination?

Big Data applications are being employed throughout the economy and society. The technologies employed are exciting, involve analysis of mind-numbing amounts of data and require fundamental rethinking as to what constitutes data. And the potential for gain through use of these technologies seems to far exceed the benefits achieved so far. However, there are issues as to what constitutes "appropriate" use of these capabilities. Societal responses to those issues, especially as it relates to privacy, will shape the future growth of Big Data.

The following discussion will tend to emphasize economic factors. However, consumer and societal forces also can materially affect technology adoption. In agribusiness, two such important forces relate to environmental and food safety concerns. If society demands (regulates) verification of agriculture's effect on an environmental phenomenon (for example, fertilizer use and its effect on the Gulf of Mexico), application of Big Data approaches may prove to be the most feasible means to respond. Similar possibilities exist relative to food safety concerns. In those instances, application of Big Data may be essential to provide "freedom to operate" in the sector.

The path by which Big Data could affect agriculture is not determined at this point. A recent article focused on identifying the "right questions" to pursue when digital technologies have the potential to disrupt industries (Wilmot 2013). The first question was, "How will IT change the basis of competition in your industry?" While we expect Big Data applications to be used within agriculture, that doesn't mean that all applications will change the basis of competition. Indeed, the question in this section's title hopefully reminds us -- there's often a thin line between vision and hallucination.

The following discussion briefly reviews the key distinctive aspects of Big Data and discusses two technology advances that can markedly change what we think of as data in agriculture. Then two innovations will be presented where Big Data is being employed to develop potential services for agriculture. Third, the issue of industry transformation will be discussed relative to technological and organizational dynamics of the sector.

Big Data's Distinctive Features

Earlier, we considered Big Data's three dimensions; Volume, Velocity, and Variety. Variety, in many ways, is the most intriguing of these dimensions. No longer is data just numbers in a spreadsheet. A number of examples have been discussed in this article. The data types included in those examples include:

- Financial transactions
- Movements of a cursor on a webpage
- "Turns of a screw" in a manufacturing process
- Tracking of web pages examined by a customer
- Photos of plants
- GPS locations
- Text
- Conversations on cell phones
- Fan speed, temperature, and humidity in a factory producing motorcycles
- Images of plant growth taken from drones or from satellites
- Questions

The capability to rapidly process large quantities of data is one necessary feature of Big Data. Analytics, "making sense" of massive amounts of highly variable data types, can provide new insights. Our history conditions us to frame analysis as a process where data is assessed to provide an answer. However, today, powerful analytical tools today can search unstructured data with the goal of "identifying the questions" embedded in the data. The NSA use of Big Data tools noted previously appears to be an example of application of this capability.

As well, continuing technology advances are fueling Big Data applications and changing the sense of what we regard as data. Let's consider two types of innovations, remote sensing and the cell phone, as potential <u>sources</u> of data which when combined with other types of data can affect the decision making capabilities of agricultural managers.

To this point, use of GPS and VRT-related tools in crop farming have been focused on input application at pre-planting and planting and yield measurement at harvest. Measurement during the growing season, either to inform input application or to assess and learn from phenomena that occur then, typically isn't done extensively for the major crops. Of course, given prior technology, the cost of conducting these measurements exceeded the perceived benefits. However, technology developments are emerging which could change this cost/benefit ratio.

A recent *Wall Street Journal* article outlined on-going efforts to transform Unmanned Aircraft Systems (UAS) capabilities originally focused on military purposes to applications supporting production agriculture. "As the spring growing season unfolds, universities already are working with agricultural groups to experiment with different types of unmanned aircraft outfitted with sensors and other technologies to measure and protect crop health". Applications include:

- Monitoring of potato production (Oregon State University),
- Targeting pesticide spraying on hillside vineyards (University of California-Davis),
- Mapping areas of nitrogen deficiency (Kansas State University), and
- Detecting airborne microbes (Virginia Polytechnic Institute and State University).

Those specific examples are only a sample of the numerous experiments and demonstrations being conducted to identify cost effective means to employ the UAS technology to enhance agricultural systems. UAS capabilities offer flexibility and potentially lower cost relative to the use of even small manned aircraft, especially for monitoring and measurement. Although many of these efforts are being done in the Midwest, it is likely that initial commercial application will occur where higher value crops dominate. Of course, an efficient process for regulatory approval of UAS flight will be needed before widespread commercial application of UAS can be implemented.

UAS innovations have the potential to make airborne monitoring more cost effective, but there are significant technical and economic issues that need to be addressed to achieve extensive viability. For some time, satellite-based sensing abilities have been available and it is believed that information from these sources has been employed to inform crop production estimates during the growing season.

Remote Sensing-Based Information and Insurance for Crops in Emerging Economics (RIICE 2013) is a global initiative currently being implemented. The RIICE consortium of partners, which includes technology, insurance and development entities, will rely upon satellites of the European Space Agency to provide on-going observations of rice production systems in Southeast Asia. The system will measure temporal changes in reflectivity of the plants to provide estimates of growth of rice plants. The Synthetic Aperture Radar (SAR) technology employed offers an effective alternative to optical observation, which has the major disadvantage that clouds can obstruct mapping and monitoring. Because satellite coverage of any specific spot on the earth occurs only once every few days, the presence of clouds when the satellite goes by can leave significant gaps in the data gathering process.

While we've become accustomed to the extensive use of cell phones in the United States, it's awe inspiring to observe cell phone adoption in developing countries. It is estimated that there were 6.4 billion mobile phone users globally in 2012 (International Telecommunication Union). Examples of the reach of cell phone use by 2012 include:

In India
- o 865 million mobile phone subscribers
- o #2 Global Market (After China)

In Brazil
- o 248 million mobile phone subscribers
- o #6 Global Market (After China, India, USA, Indonesia, and Russia)

In Nigeria
- o 113 million mobile phone subscribers
- o #10 Global Market

Adoption is extensive in urban areas; however, cell phone use is common in rural areas. Cell phones, of course, are not created equal, ranging from devices which receive phone and text messages to smart phones that effectively operate as small computers. In addition to receiving data and information, the use of cell phones as sensors is receiving intense interest from both private and public sector entities. The camera is a means of capturing images, which new technology is allowing to be analyzed as digital data.

Messages sent by cell phone users also can provide important data -- that when aggregated can provide important and timely insights. Figure 1 illustrates this point (Mock et al. 2013). A key factor indicating social well-being in developing countries is the food price index. While extensive efforts are made to track food prices, official reporting processes take time to collect data and therefore may unduly lag actual conditions affecting low income consumers and social stability. However, individuals "talk" about changes in food prices continually. The two graphs in Figure 1 both track movements in food prices in Indonesia during 2010 and 2011. The bottom graph provides the official monthly inflation rate for food prices. The top graph tracks the volume of tweets per month in Indonesia, where the topic was the price of rice. The similarity in direction and turning points of the two graphs provides support for the belief that important information can be acquired from social media sites.

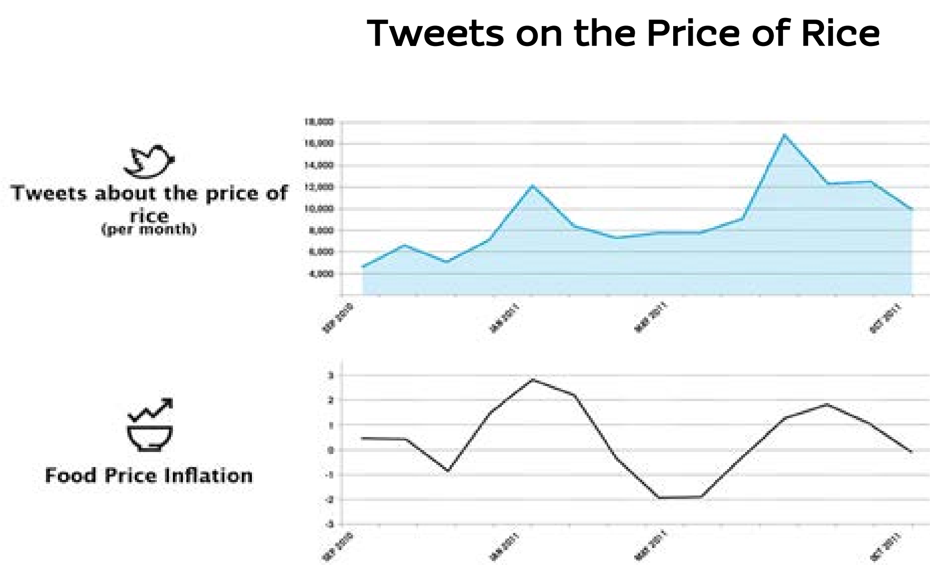

Figure 1. Comparison of estimates of food price inflation and tweets about the price of rice in Indonesia.

Two Specific Big Data Applications for Agriculture

Figure 2 displays a rendition of the Lettuce Bot, a technology being developed to identify and then eliminate weeds in the field (Dawson 2013). Developed at a California start-up, Blue River Technology, the key Big Data aspect of the Lettuce Bot is its ability to identify plants and weeds instantaneously from a database of millions of images of plants. The current version of Lettuce Bot releases a spray of fertilizer on either a weed or an unwanted plant (for example, a plant growing too close to another plant). Later versions of the Lettuce Bot may use mechanical devices to remove the offending plant, for example, in organic fields where fertilizer application would not be appropriate.

Figure 2. Depiction of the Lettuce Bot weeding a vegetable field.

The Lettuce Bot has the potential to reduce or eliminate hand-weeding or thinning practices. In labor-intensive vegetable production systems, access to labor and compliance with labor regulations are significant managerial issues. Similar applications of machine recognition tied to image-based databases could be applied at several steps in agricultural supply chains, especially where use of manual labor to do sorting is an issue.

All of us are familiar with general weather forecasts delivered by the media. However, most of us are frustrated when such forecasts don't meet our business needs. Figure 3 describes Deep Thunder, an initiative from IBM, which focuses computing power, multiple data sources, and targeted software to provide "hyperlocal" weather forecasts. Here, hyperlocal refers both to geographic scale and to specificity of business needs.

Figure 3 cites numerous potential benefits of use of Deep Thunder-based services. These include direct linkage to precision agricultural practices, increased yields, reductions of postharvest loss, and consumer benefits of lower price, improved quality and presumably lesser environmental impacts. A key factor noted is the potential for more effective water use, a critical global concern.

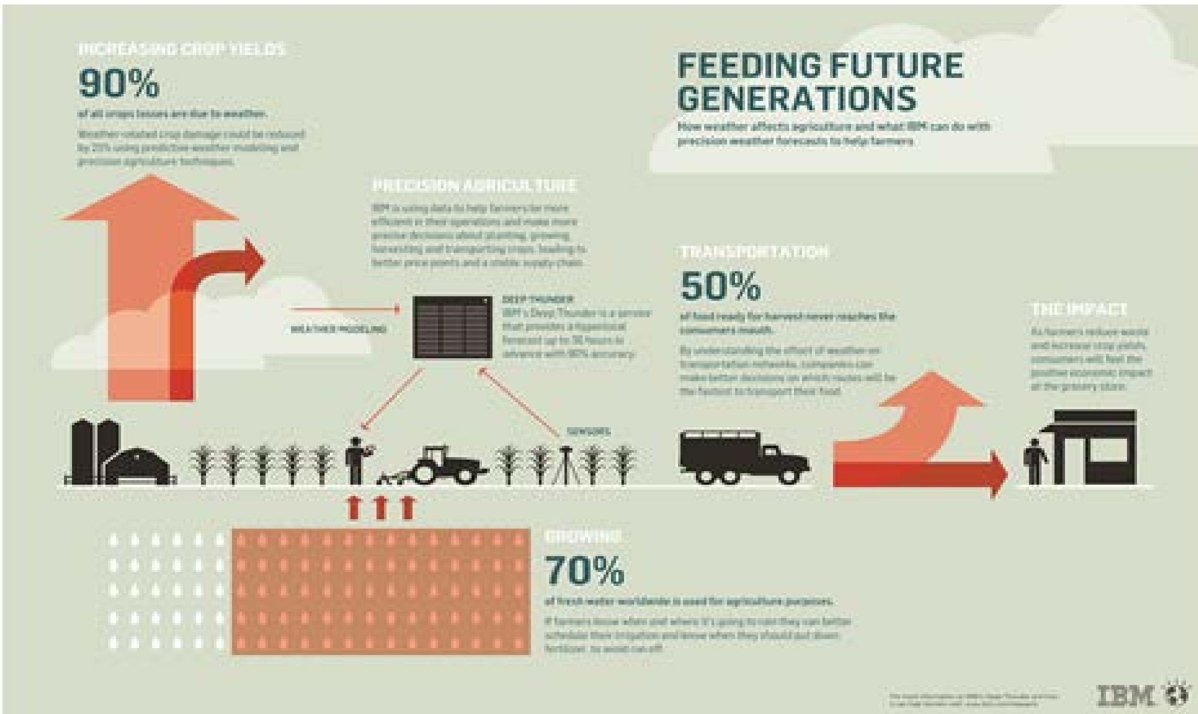

Figure 3. Depiction of use of Big Data technology to provide hyperlocal weather forecasts to support agriculture.

Speculation about Changing the Basis of Competition

Important in their own right, the Lettuce Bot and Deep Thunder applications also were shown to illustrate a specific type of application of Big Data technology in agriculture. In these instances, the Big Data elements of the application are available independently of the agribusiness operations to which they might be applied. Individual managers can assess the value of the offered service and then adopt, or not, the service. The Big Data basis of the service is essentially irrelevant to the manager making the purchase decision. Such innovations will affect competition in specific markets, but are not likely to alter the basis of industry competition. Alternatively here are three examples of altered competition:

- Wal-Mart transformed retailing through an aggressive focus on price facilitated through path-breaking use of IT and by using those capabilities to alter relationships with suppliers.
- Amazon enhanced, in some dimensions, the customer's shopping experience and employed ICT to learn how to improve each customer's next experience.

- Toyota's lean manufacturing techniques are based upon extensive data analysis leading to transformed relationships with suppliers.

Cutting across these and other examples, common features of changes in the basis of competition include:

- Dramatic cost reductions.
- Quality enhancements desired by customers.
- Redefined relationships across stages of the value chain.

The most dramatic and impactful of these instances results when the features occur in combination.

Figure 4 provides a highly useful illustration of data integration and analysis as an attractive vision for production agriculture (Riverside Research). In several respects the perspective shown there is similar to our now familiar concepts of precision agriculture as they affect field operations and harvest. But there is one important point of difference, the explicit emphasis on analytics and integration. Indeed Figure 4 explicitly draws attention to the need to integrate across many production sites rather than to expect to make substantial advance based upon precise data from a single field.

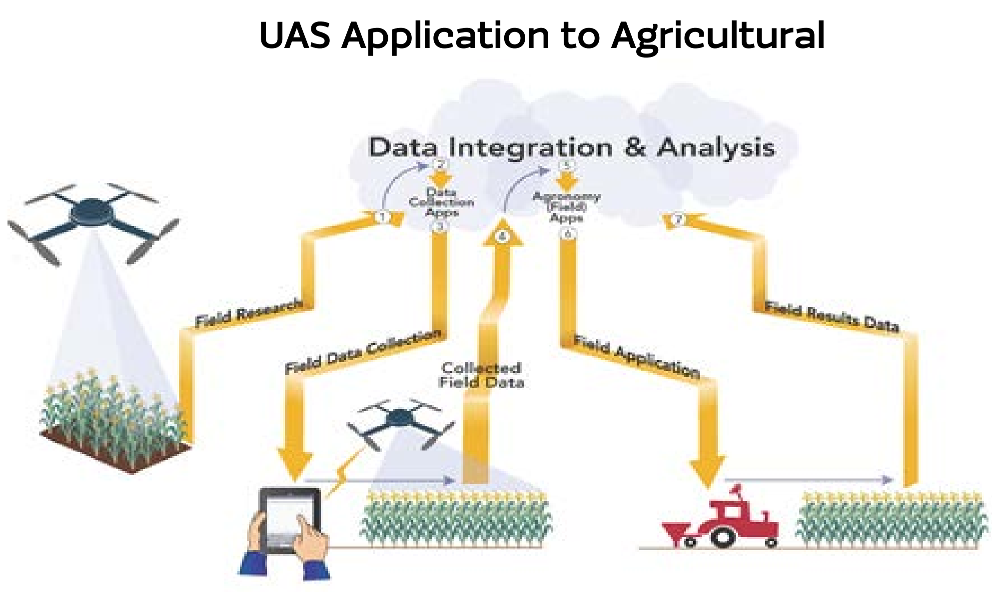

Figure 4. Representation of data integration and analysis for production agriculture[1]

Figure 4, although useful, masks two key factors that will have to be addressed if Big Data applications are to alter the basis of competition in agriculture. One factor relates to the word integration, in particular, the process of integrating across numerous sources of data – where the control of those data doesn't exist within a single entity. The second factor relates to the word

[1] Courtesy Riverside Research, Champaign, Illinois.

analysis. Figure 4 suggests that the needed data can be obtained from field operations, just as counting the turns of a screw can determine if manufacturing processes are within bounds. Field data, while essential, likely is not sufficient to truly optimize production agriculture.

Let's address each of these two factors, noting that the aggregation issue has two dimensions;

- Would an individual farming operation have enough data and/or, more importantly, the analytic capacity to create substantial new knowledge?
- How could the relevant data captured at numerous points across the agricultural value chain be aggregated and then employed to enhance operations both on and off the farm?

Commercial farms continue to grow in size, but the individual farm likely won't have sufficient scale and variety of operation to create significant new information to optimize operations. Indeed, what the individual farm operator doesn't have is detailed information on other similar units and the effect of variations in practice between farms. Scale also is critically important to justify the cost of analytic capabilities. Optimal use of field data likely will require aggregation across a number of farming units. The issue of scale also is a factor in animal agriculture.

Also, the farm operator may not have all the potentially relevant data that relates to the operations of the farm. VRT application of inputs often is done by agricultural retailers. Although planting is most typically done by the producer, the genetic capabilities of the seed planted, are not captured simply by knowing variety planted. Remote sensing, via UAS or satellites, likely will not be performed by individual farmers. Remote capture of engine and machine performance likely is best done at the manufacturer level. Therefore optimal aggregation would need to be accomplished among a number of firms even if some of the firms are competitors.

The second factor relates to the nature of the knowledge required to optimize agricultural performance. Figure 4 focuses on operational data from the farm field. A similar depiction for animal agriculture would depict the feedlot, dairy barn, farrowing and finishing unit, etc. However, having even comprehensive information based upon operational performance may not be sufficient to optimize these production systems.

Agricultural production exploits biological processes, where key factors (such as weather events and pest infestations) are not controlled – even if they are measured. US and global agriculture has made tremendous productivity advances in the last 100 years through the application of what we might call, "Small Data". Knowledge of science and engineering was necessary to unravel the interactions of biology, in the context of those uncontrollable factors. 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.

An important aggregation question is; how can the best aspects of the Small Data system be linked to the application of Big Data technologies? This factor is particularly intriguing in the context of the on-going evolution of agriculture. Not only do weather events vary from year to year but pests evolve in their location and behavior. Advances in genetics and research continue

to provide enhanced but differing capabilities. Therefore it seems unlikely that even the most comprehensive data from operations will be totally satisfactory in agricultural systems.
Big Data enthusiasts may not appreciate this element of agricultural systems. A recent book (Mayer-Schonberger and Cukier 2013) on Big Data asserts:

> *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*.

Knowing, at increasing levels of precision, "what" happened in the field or in animal facilities does have value. However, not also knowing "why" is a concern in agricultural applications because at least some of the key factors that led to this season's "what" likely will not be identical for the next season or production cycle.

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, those functions largely shifted to the private sector in the United States. It is possible that effective linkage of these differ types of information also will occur in the private sector in regions of the world where commercial market channels are strong. It is less clear how the capabilities that link traditional ag science with data from the on-going operations of production agriculture would occur in developing agriculture settings.

From a pure technological capability perspective, the aggregation challenges just noted are not insurmountable. However, the associated economic and organizational issues likely will be significantly more difficult to surmount:

- The potential net economic benefits have to be significantly positive to initiate information system innovation. Often where ICT has made significant impact, the information capture activity was done at low or no cost. For each transaction, the cost of capturing the customer's loyalty card information approaches zero. Therefore extracting value by aggregating information that comes at such very low cost is attractive. Or if the missile may explode (and financial penalties imposed on the manufacturer) if screws aren't affixed in exactly the correct manner, the economic justification of the cost of monitoring how screws are affixed is straightforward.

- The organizational challenges of achieving "between firm collaboration" are significant. Such challenges have significant economic implications when the potential parties are competitors or when they are suppliers and customers. Relative to Big Data, the potential relationships would be affected by the different types of capabilities required to achieve advances in knowledge. The analytics required will involve very specialized skills and there is little reason for each party to possess those capabilities. However, the potential for "hold-up" would be a concern of the firms that don't possess that capability.

Of course, producers have a tradition of using the cooperative model to achieve collaboration between firms. In several Midwest states, farm records systems operate to achieve similar goals to those noted here. Where collaboration at the input supply and the producer level is required, effective incentives structures (which may be both financial and non-financial in nature) will need to be implemented.

In addition to organizational challenges, implementation of information systems generally are characterized by relatively large initial investment and relatively low operating costs. These characteristics typically lead to first-move, winner-take-all effects (Shapiro and Varian 1999). As noted previously, at one time public funding and Land Grant/USDA institutions might have been expected to conduct the necessary research and development. Today, at least in the United States, that seems unlikely. Private firms, both those operating in agriculture and in the technology sector, are exploring Big Data opportunities. Required key capabilities, however, aren't likely to totally exist within any one of the existing challengers in either sector suggesting a fertile field for experimentation and innovation.

Wrapping It Up

Agriculture, globally and in the United States, has been identified as a target for Big Data application by technology developers, including both startups and multinationals. The established firms include those currently in the ag sector or those in the non-ag technology sector.

While the term Big Data is new, application of ICT-based innovation has driven economic transformation over the last three decades. In the 1990s, "the knowledge economy" was the hyped term of the decade and understanding the transformative role of ICT was a key research and practical question. Findings from that work focused on the need for implementation of low-cost means to capture numerous information attributes at the time transactions occurred. Although low-cost separability is necessary, industry transformation occurs only when the information acquired can be aggregated to form new knowledge --which leads to novel operations and offerings. Although the terminology has changed, these findings also apply to the successful application of Big Data technologies today.

The three commonly cited characteristics of Big Data are Volume, Velocity and Variety. A key factor to understanding the potential of Big Data is to realize that it is not **just** about lots of numbers. The variety characteristic emphasizes that data now includes a stunning range of phenomena. Further, it is important to appreciate the power of "analytics", where findings and insights are gained from multiple data sources that differ in structure and original purpose.

The experience of successful Big Data application in non-ag sectors uniformly stresses that it's the business issues and opportunities, not technological capabilities, which determine success. That counsel seems appropriate for agricultural managers as well. Existing technologies, and those on the near horizon, suggest considerable potential for Big Data applications in the agribusiness value chain. Some of those applications will employ data that is not directly captured from agricultural operations to provide enhanced product and service offerings to the sector. These offerings will be adopted or not based upon well accepted cost/benefit parameters.

The more intriguing potential relates to Big Data applications where a considerable segment of that data is generated from within the sector's operations. Data aggregation across business entities seems to be a necessary component of these applications. Competitive dynamics and intellectual property concerns will join expected net benefits among the several factors that will be needed to be addressed. For many Big Data applications, valuable business insights can be gained from capture and analysis of operational data only. However, the biologic underpinnings of agricultural production limit the gains that can be gathered solely from operational data. Linking data from operations with information and knowledge from laboratories and experimental sites (ag's Small Data) will be required to effectively optimize sector performance.

Over the last 50 years computation and analysis have enhanced performance in the economy and in agriculture. Ag sector analysts contributed to those enhancements. However with the limited computer power available then, a key to success was to effectively constrain the problem to fit the data and computational power available. No longer is that constraint needed as Big Data approaches come to agribusiness. Successful application will occur in US and global agriculture but that success will be determined as much by organizational and managerial factors as by technology.

References

Barton, Dominic and David Court. Making Advanced Analytics Work for You. 2012. *Harvard Business Review* October.

G. Dawson. Lettuce Bot: Roomba for Weeds. Modern Farmer. http://modernfarmer.com/2013/05/lettuce-bot-roomba-for-weeds/

Economist Intelligence Unit. 2012. Big Data: Lessons from the Leaders.

Hagerty, J.R. How many turns in a screw? Big data knows. 2013. *Wall Street Journal* May 15. Online: http://search.proquest.com/docview/1351390761?accountid=14553.

Hickins, Michael. How the NSA Could Get So Smart So Fast. 2013. *Wall Street Journal* June 12. http://online.wsj.com/article/SB10001424127887324049504578541271020665666 .html?mod=wsj_ciohome_midLatest.

Kirkland, Rik. Connecting Everything: A conversation with Cisco's Padmasree Warrior. http://www.mckinsey.com/Insights/High_Tech_Telecoms_Internet/Connecting_everything_A_conversation_with_Ciscos_Padmasree_Warrior?cid=disruptive_tech-eml-alt-mip-mck-oth-1305.

Lu, J. and M. Bowles. 2013. How Will Nanotechnology Affect Agricultural Supply Chains? *International Food and Agribusiness Management Review* 16 (2): 21-42.

V. Mayer-Schonberger and K. Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work and Think.* New York. Houghton Mifflin Harcourt Publishing Company.

McAfee, Andrew and Erik Brynjolfsson. 2012. Big Data: The Management Revolution. *Harvard Business Review*. October.

Sonka, Steve. 2013. Beyond Precision Agriculture; If Big Data's the Answer, What's the Question?[2] Presented at the 2013 Midwest Food and Agribusiness Executive Seminar. Center for Food and Agribusiness Center, Purdue University.

McKinsey Global Institute. 2011. *Big Data: The next frontier for innovation, competition, and productivity*. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation May.

Mock, N., N. Morrow and A. Papendieck. 2013. From Complexity to Food Security Decision-Support. *Global Food Security* 2(1):41-49. http://dx.doi.org/10.1016/j.gfs.2012.11.007

Sampler, J.L. 1998. Redefining Industry Structure for the Information Age. *Strategic Management Journal*. 19(4)343-355.

Shapiro, C. and H.R. Varian. Information Rules. 1999.

P. Wilmot. 2013. The do-or-die questions boards should ask about technology. McKinsey & Co. June. http://www.mckinsey.com/Insights/Business_Technology/The_do-or-die_questions_boards_should_ask_about_technology?cid=other-eml-alt-mip-mck-oth-1306

About RIICE. http://www.riice.org/about-riice/

Clifford, S. and Q. Hardy. 2013. Shoppers: Store Is Tracking Your Cell. *The New York Times* July 14. Online: http://www.nytimes.com/2013/07/15/business/attention-shopper-stores-are-tracking-your-cell.html?pagewanted=all&_r=0

International Telecommunication Union. 2013. Key ICT indicators for developed and developing countries and the world. Online: http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2013/ITU_Key_2005-2013_ICT_data.xls

International Telecommunication Union. 2013. Mobile-cellular subscriptions. Retrieved from: http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2013/Mobile_cellular_2000-2012.xls

Wikipedia. *Big Data*. http://en.wikipedia.org/wiki/Big_data [accessed June 28, 2013].

[2] Based upon a McKinsey & Company case study *Beyond Precision Agriculture; If Big Data's the Answer What's the Question?* 2013.

