PURDUE UNIVERSITY

Center for Food and Agricultural Business

CAB CS 13.2

September 1, 2013

Beyond Precision Ag: If Big Data's the Answer, What's the Question?

Introduction

Remember the Midwest farm sector of 1993 — just 20 years ago? While much is similar to today (large combines, large tractors and large fields), a number of technologies we take for granted today did not exist or weren't commonly used only two decades ago. Who would have thought that a farmer wouldn't have to actively steer the tractor or combine as it moved across the field? Or that, while sitting in the combine, the farmer primarily would be paying attention to the readout on something called a yield monitor? Or maybe, the farmer would be using a cell phone (not a CB radio) to talk to a market advisor or to "surf the Web?" Over the last 20 years, information and communication technology (ICT) innovations have formed the basis of the tools that we now call precision farming. Use of these tools has changed farming operations and enhanced operational efficiency.

Of course, ICT application has not been limited to agriculture. What has happened in other economic sectors over that same two decades? In the early 1990s, people enjoyed watching movies at home and only needed to stop at the video store to rent a movie. To enjoy reading a book, one could stop by the local Borders bookstore to select an attractive title and maybe have a cup of coffee. If the consumer wanted to enjoy a few days of travel, one could call a travel agent and receive an airline ticket (printed on a real piece of paper!).

But, how would that consumer decide which movie, book or destination they would most enjoy? Couldn't they just "Google it?" Oh wait — Google, Amazon and the Web weren't part of our popular lexicon, and they certainly weren't tools used daily by the consumer — or by the typical agribusiness manager. In many sectors, advances in the use of ICT have fundamentally transformed the way business is done — they changed the basis of competition.

Today, agricultural managers are hearing about additional exciting opportunities fueled by ICT-based advances. The one that is receiving the most attention (for the changes it will cause

This case study was prepared by Steve Sonka, director of the ADM Institute for the Prevention of Postharvest Loss and emeritus professor of agricultural strategy at the University of Illinois, as a basis for class discussion and represents the views of the author, not the university. The author would like to thank the numerous individuals from agribusiness, technology and academia who provided counsel and guidance in the development of the case study.

The First Tell

I recently attended a presentation entitled, "Surprise is inevitable, Being unprepared is not". In 1995, when Jeff Bezos started Amazon.com in his garage to sell books on the Internet, it's unlikely that senior managers at Walmart, Sears or Target paid too much attention. Books were an insignificant part of their sales and they certainly didn't fool around with that Internet thing. Yet today major retailers rely on Internet-based sales.

Big Data applications today aren't focused on agriculture and the food chain. Some of us likely will be surprised at the impact Big Data does have in the future. Others of us likely will be surprised if it doesn't have major impact. Either way, we should be prepared.

- Exhortations focused on the virtues and evils of Big Data are almost impossible to ignore or escape.
  - The technologies and the potential capabilities are indeed noteworthy and, given what we currently know, will have significant impact on the economy and on many firms.
  - As is typically the case, we're likely to overestimate the immediate effect and underestimate the longer term impact.
- Big Data refers to more than the volume of data.
  - Velocity and variety of data are additional key factors.
  - Variety may be the source of the biggest managerial surprises — the movement of your eyeballs as you read this sentence could be data.
  - Analytics refers to the discovery of meaningful patterns in data.
  - The combination of volume, velocity and variety fueled by sophisticated analytics is the source of Big Data's power.
- The tools and techniques of precision agriculture are likely to be essential components of future Big Data applications that involve production agriculture, its suppliers and its customers.
- While it is natural to become focused on the possibilities of new technologies, basic economic and business factors will drive successful adoption. Among these are:
  - Low-cost data acquisition and the ability to create new value from aggregating data typically characterize successful initiatives.
  - The nature of information technology-based innovations leads to winner-take-all effects.
  - Changes in regulatory and consumer sentiment can turn strategic options into a cost of doing business.

Exhibit 1. Tell them what you're going to tell them

in business, government and society) is called "Big Data," a term often followed by the phrase, "whatever that is." There is considerable recognition that Big Data is relevant for agriculture. Indeed, Padmasree Warrior, chief technology and strategy officer for Cisco Systems, 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 that use water, light and sunlight more efficiently.[1]

This case study explores the potential for Big Data in agriculture. In doing so, we'll examine the question posed in the title: "If Big Data's the answer, what's the question?" One question might be, "Could application of Big Data fundamentally transform (change the basis of competition in) Midwestern agriculture?"

The remainder of this case study is comprised of the following components:

- Precision agriculture, its effects and emerging applications of ICT
- CT application and industry transformation
- Big Data and its business application
- Big Data and Midwest agriculture: vision and/or hallucination?
- Wrapping it up

All of us have probably heard the advice that a good speaker needs to "tell, tell, tell" an audience. In that spirit, Exhibit 1 summarizes the "tell" for the following text.

Precision Agriculture, Its Effects and Emerging Applications of ICT

You Can't Manage What You Don't Measure!

Attributed to both Peter Drucker and W. Edwards Deming, this phrase is as applicable to farmers as it is to managers at Caterpillar or Boeing. Growing up on an Iowa farm, this practice was introduced to me at an early age. In those days, we had to carry the (hopefully full) milking machine from the cow to the milk tank and it was fairly easy to know which cows were producing more. Because there were fewer than 20 cows, it was also possible to give an extra scoop or two scoops of grain to those high producers. In the same vein, it was also possible to put more fertilizer on fields with higher-productivity ground. However, the linkage between the management action (more grain, more fertilizer) and the resulting economic productivity was imprecise at best.

The desire to link outcomes and management actions in farming is not a new phenomenon. However, the economics of measurement (the cost of measurement versus the benefits of doing

---

[1] 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 McKinsey&Company, 2013.

so, given the available technology), inhibited my dad and farmers like him from being more precise. Therefore, in the early 1990s, when technologies associated with what we now call precision agriculture (or precision farming) started to become available, there was a receptive, but appropriately skeptical, audience among Midwestern farmers.

This section of the paper will provide an overview of our precision farming experience. It will identify popular technologies employed and will review available evidence as to the economic value of precision farming. Emerging technologies also will be discussed. This section is not intended to be a comprehensive assessment. Rather, it is meant to inform the following discussion of the potential for Big Data in Midwest agriculture. The current tools and techniques of precision farming have existed largely without Big Data concepts. However, it is likely that Big Data efforts involving production agriculture will need to employ precision agriculture practices.

Twenty Years of Precision Agriculture

Although the term "precision agriculture" is now familiar, a specific, common definition does not exist. A 1997 report of the National Research Council (NRC)[2] refers to precision agriculture as a "management strategy that uses information technologies to bring data from multiple sources to bear on decisions associated with crop production." Key technologies and practices noted include:

- Geo-referenced information
- Global positioning systems
- Geographic information systems and mapping software
- Yield monitoring and mapping
- Variable-rate input application technologies
- Remote and ground-based sensors
- Crop production modeling and decision support systems
- Electronic communications

The technologies and practices noted in that list are applicable to both crop and livestock production, where georeferencing can refer to both sub areas of a field and individual animals. The tracking processes and required tools may differ but the managerial goal is still to be able to separately manage increasingly smaller units.

Since 1997, technologies have advanced, although the general categories remain relevant. For example, autosteer capabilities on farm machinery have become much more prevalent over the last decade. And active measurement of the planting process (recording where "skips" occur) is now feasible. Further, the ability exists to sense the status of farm machinery and to electronically communicate when operations are out of acceptable bounds.

[2] Board on Agriculture, National Research Council. Precision Agriculture in the 21st Century. 1997.

In the mid-1990s, a group of agribusiness professionals in Champaign County, Ill., came together to explore the opportunities associated with two emerging technologies, precision agriculture and the Internet. This group, called CCNetAg, was part of an initiative co-sponsored by the local Chamber of Commerce and the National Center for Supercomputing Applications at Illinois. A voluntary enterprise, CCNetAg provided a vehicle for farmers, managers and researchers to jointly explore the adoption of these tools. One of the group's activities was to create a "picture" of a future where precision agriculture and the Internet were effectively joined (Figure 1)[3]. Although it was created some time ago, the graphic continues to depict key elements of precision farming:

- The role of geo-referencing is depicted by the satellites with linkages to the farm field.
- On the field itself, key farming operations are capturing and being directed by digital information:
    - Soil sampling
    - Nutrient application
    - Planting
    - Crop scouting
    - Harvesting
- The layers that underlie the farm field represent the notion that visual mapping would allow the farmer, and the farmer's advisors, to see meaningful correlations that would then inform future decisions.

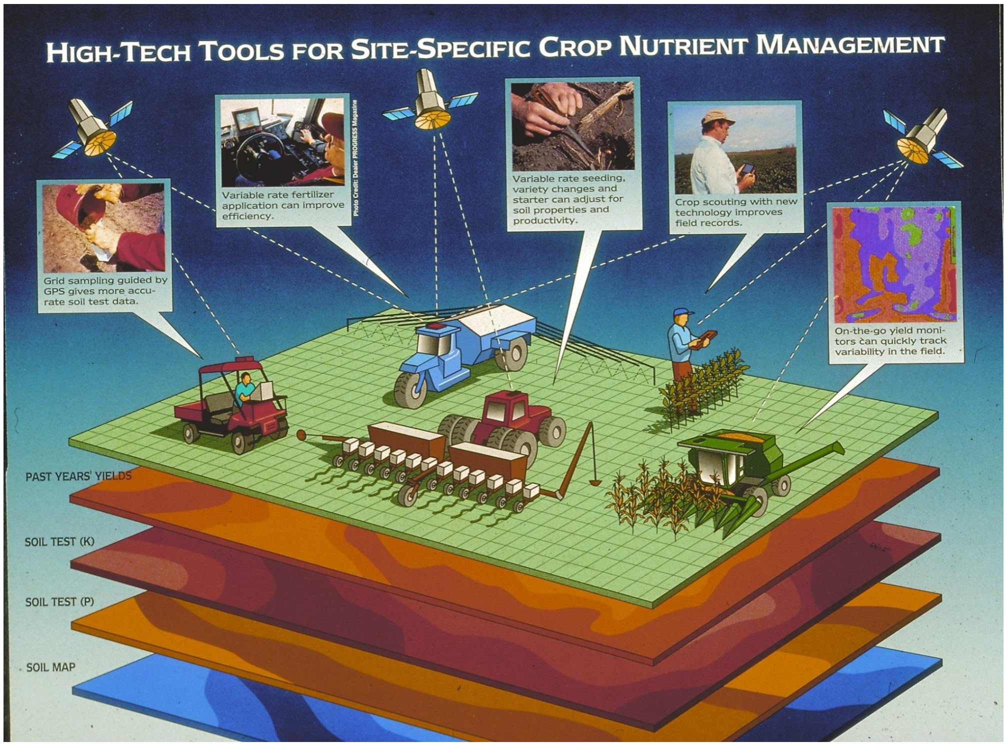

Figure 1. A mid-1990s view of precision farming from the CCNetAg group

[3] Courtesy of Harold Reetz, Monticello, Ill.

Figure 1 also is informative because of deficiencies that have now become apparent. For example,

- Visual analysis across maps denoting several differing factors initially seemed promising, but has been less informative than desired. The effect of single factors, such as a blocked tile line or a buried fence row, sometimes becomes visible. However, trying to discern the yield effects of soil type, several nutrients and seed variety solely by comparing individual maps of each of the factors has been difficult.

- The technologies and tools employed are subject to error, especially when operating in a dirty, mobile and time-pressured environment.

- The extent of measurement precision differs markedly between "levels." In particular, information as to the underlying soil capabilities exists at a much more general geographic scale than does information regarding other practices.

- The graphic of Figure 1 concentrates on the farm field. However, farming is conducted across multiple fields and within the context of extensive food supply chains. Advances in efficiency that can be gained by enhancing coordination between fields and firms are outside the graphic's vision.

The fundamental concepts of Figure 1 have analogous features in animal agriculture. While advances have occurred throughout animal agriculture, the dairy sector is of particular interest as technologies that focus on the individual dairy cow have come into play. For many years, it has been possible to automatically track production per cow. Information on milk quality, measured at the processing unit, has been rapidly transmitted to the farmer. Today, technological advances are linking observation of those quality factors to the individual cow. Devices to continuously monitor the health characteristics of the animal, including indicating when artificial insemination should take place, are becoming increasingly available. Technologies to allow robotic milking have become entrenched in the industry. While direct economic effects are important, benefits associated with the robotic milker include expanding the available labor supply and enhancing quality of life for the farm family.

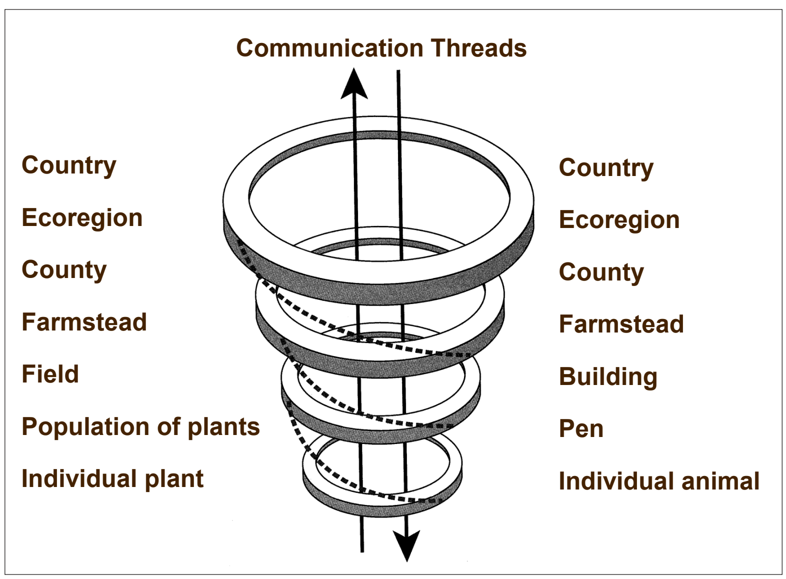

Figure 2. A conceptual view of value enhancement as information is aggregated

Focusing on that broader perspective, the previously noted NRC report[4] proposed a communication spiral (Figure 2) that depicts the theoretical goal of maximizing effectiveness within the sector. A fundamental notion of this figure is that the value of precision farming data is enhanced as it is aggregated and assessed. Whether focused on crops (on the left) or on animal agriculture (on the right), value (economic, environmental and/or social) increases as the level of aggregation expands. The growing size of each spiral denotes that gain in value.

Use and Effect of Precision Agriculture in the Midwest

Although anecdotes abound, little research information is available documenting the use and economic effect of precision agriculture in the Midwest. In this section, we'll examine available information regarding farmer adoption (using both USDA and Ohio data) and insights from farm input suppliers.

In 2011, the Economic Research Service of the USDA published an analysis focused on precision agriculture adoption.[5] The Agricultural Resource Management System (ARMS) survey provided the data for this experiment. Although an ARMS survey takes place each year, the survey does not collect data for each commodity in every year.

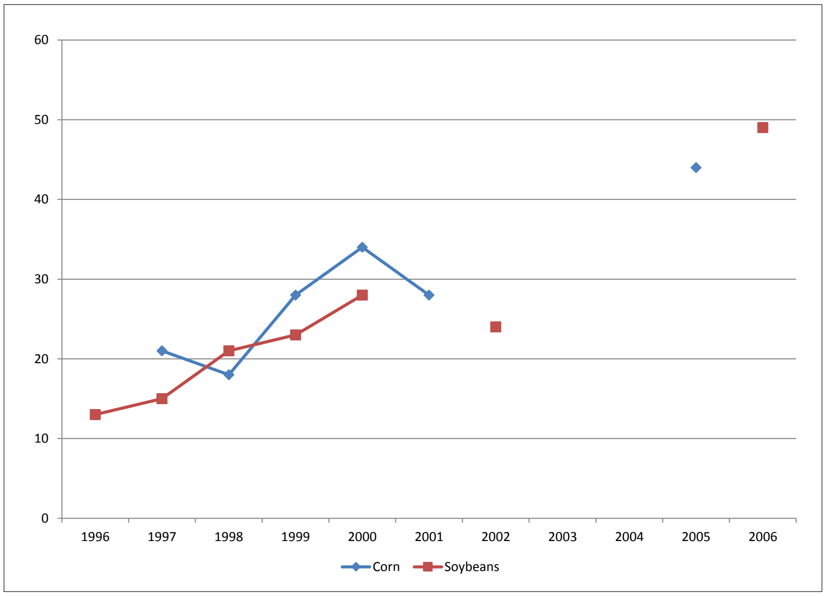

Figure 3. Yield monitor adoption in the Corn Belt; 1996-2006

Figure 3 shows the extent of yield monitor adoption reported by Midwest farmers in the ARMS survey from 1996 to 2006. Not surprisingly, the survey shows a rising trend in adoption for both corn and soybeans with levels exceeding 40 percent by 2006. The use of variable-rate technology allows for the tailoring of input application (nutrients and seed) to differing conditions within a specific field. Figure 4 illustrates the reported use of variable-rate technology as being relatively flat, at around 10 percent to 12 percent for corn and 8 percent for soybeans, in the late 1990s and early 2000s.

---

[4] Adapted from Board on Agriculture, National Research Council. Precision Agriculture in the 21st Century. 1997.
[5] Schimmelpfennig, D. and R. Ebel. On the Doorstep of the Information Age. Economic Information Bulletin Number 80. USDA. August 2011.

Beyond indicating rates of farmer adoption, the USDA report compares performance characteristics of adopting farmers with of those non-adopting farmers. Key insights gained from that correlation include:

- Corn and soybean yields were significantly higher for yield monitor adopters than for non-adopters nationally. The yield differential for corn grew from 2001 to 2005. Yield monitors are being adopted more quickly by farmers who practice conservation tillage.

- Corn and soybean farmers using yield monitors had lower per-acre fuel expenses. Average per-acre fertilizer expenses were slightly higher for corn farmers, but were lower for soybean farmers.

- Average per-acre fuel expenses were lower for farmers using variable-rate technologies for corn and soybean fertilizer application, as were soybean fuel expenses for guidance system adopters.

- Adopters of GPS mapping and variable-rate fertilizer equipment had higher corn and soybean yields.

Of course, these correlation findings don't necessarily prove causation. It may be that better farmers earn higher yields, have lower costs and employ precision farming techniques. Such inter-correlations always make it difficult to precisely measure the effect of specific management practices. Also, the mid-2000s experienced lower output prices and input costs relative to those of the last five years. Higher levels for both output prices and input prices may have fueled adoption of precision farming techniques significantly since the mid-2000s.

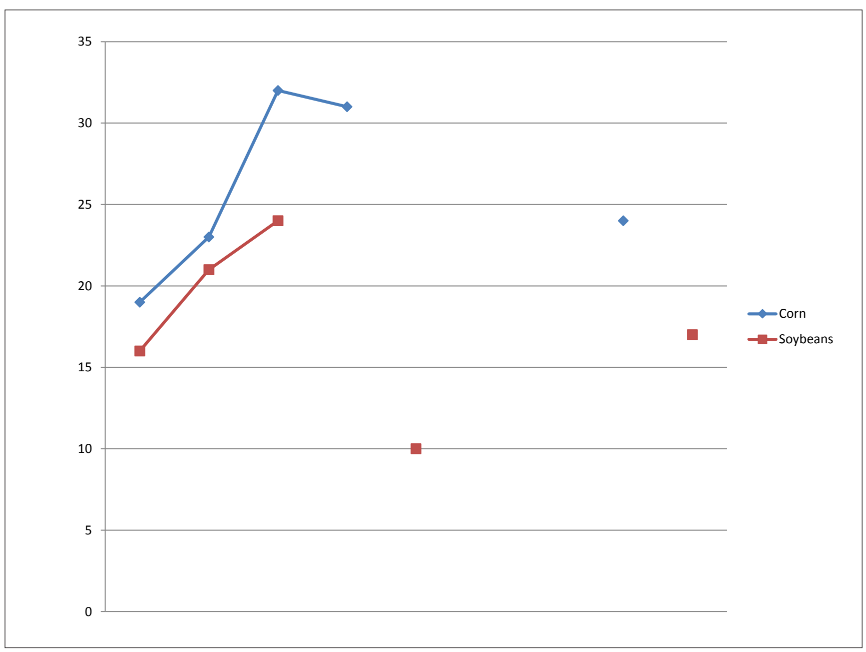

Figure 4. Variable-rate technology adoption in the Corn Belt; 1998-2006

A more recent study took place in Ohio in 2010.[6] Its findings are based on a

[6] Diekmann, F. and M.T. Batte. 2010 Ohio Farming Practices Survey: Adoption and Use of Precision Farming Technology in Ohio. AEDE-RP-0129-1. Ohio State University. Aug. 2010.

representative sample of more than 1,100 Ohio farmers, of which almost 40 percent had adopted at least one precision agriculture practice. Adopting farmers tended to have more formal education, were younger and operated larger farm units. Figure 5 highlights some of the study results regarding adoption of specific technologies. The three columns shown provide the average percent adoption for all the farms with more than 1,100 acres and the largest two farm-size categories after. The markedly higher pace of adoption for the larger farm categories is revealing. As farm units continue to consolidate, the practices of precision agriculture are likely to be even more prevalent. GPS technology, precision guidance, yield monitors and geo-referenced soil sampling are employed by the majority of Ohio's larger farms. It also is interesting that aerial/satellite field imaging is being used on about one in five of those operations.

Item		Gross farm sales	
	All farms	$500,000 to 999,999	$1,000,000 or more
		Percent	
GPS technology	30.2	58.1	78.5
Precision guidance	27.4	55.9	77.7
Yield monitor	25.3	59.1	79.7
Geo-referenced soil sampling	22.7	43.4	55.8
Variable rate application for fertilizer – potassium	19.4	36.9	49.0
Variable rate application for fertilizer – nitrogen	5.7	10.4	15.1
Aerial/satellite field imaging	9.8	19.4	20.7
Variable rate application for other farm inputs – seeds	9.0	15.4	25.5
Farmers adopting one or more of the above	38.7	68.2	84.7

Figure 5. Adoption of precision agriculture practices in Ohio

"Overall, how would you rate the costs versus benefits of this technology?"		
	Benefits significantly greater (1)	Benefits slightly greater (2)
	Percent	
VRT for fertilizer	53.8	35.7
Geo-referenced soil sampling	51.1	34.9
Precision guidance	40.7	39.5
Yield monitor	32.5	37.1
Map-based field scouting for weeds	2.6	12.8
VRT for other crop inputs – seeds	4.2	9.1
Aerial/satellite field imaging	3.7	11.7

Figure 6. Perceived net benefits of precision agriculture practices

Relative to benefits, the perceptions of the 450 Ohio producers who've adopted at least one precision agriculture practice are shown in Figure 6. The percentage of those farmers who thought each technology provided benefits were significantly or slightly greater than its costs are detailed there. A majority of adopters believe that benefit/cost ratios were significantly positive for variable-rate application of fertilizer and geo-referenced soil sampling. One-third or more saw significant benefit for use of yield monitors and precision guidance.

For more than a decade, the Center for Food and Agricultural Business at Purdue University[7] has independently, and then in combination with CropLife,[8] conducted surveys of agricultural

[7] Whipker, L.D. and J.T. Akridge. 2009 Precision Agricultural Services. Center for Food and Agricultural Business at Purdue University. September, 2009.
[8] Erickson, B., D. Widmar, and J. Holland. Survey: An Inside Look at Precision Agriculture in 2013. http://www.croplife.com/article/34108/1/survey-an-inside-look-at-precision-agriculture-in-2013

input suppliers regarding the precision agricultural services they provide. As shown in Figure 7, Midwest agricultural input dealers have employed precision agricultural tools for much of the last decade. For most of the decade, 6 out of 10 dealers used GPS guidance for input application. In recent years, the use of GPS guidance for field mapping and for vehicle logistics doubled to about 4 in 10 dealers. Now, approximately 2 in 10 dealers employ satellite/aerial imagery services — this level has also recently doubled. Although not shown in Figure 7, the two most recent surveys indicate that about 15 percent of input suppliers do not use any of the precision agriculture services tracked in the study.

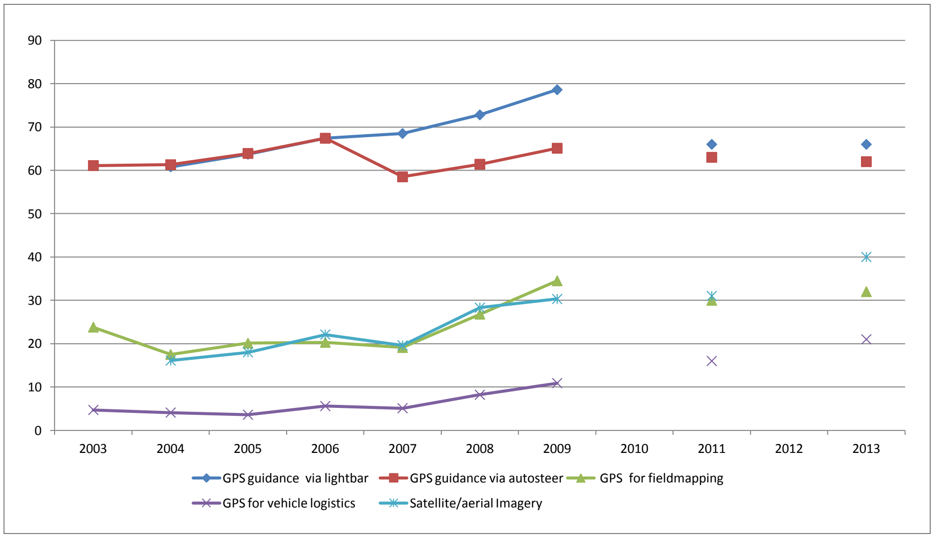

Figure 7. Input supply dealers' use of precision agricultural techniques in the Midwest; 2003-2013

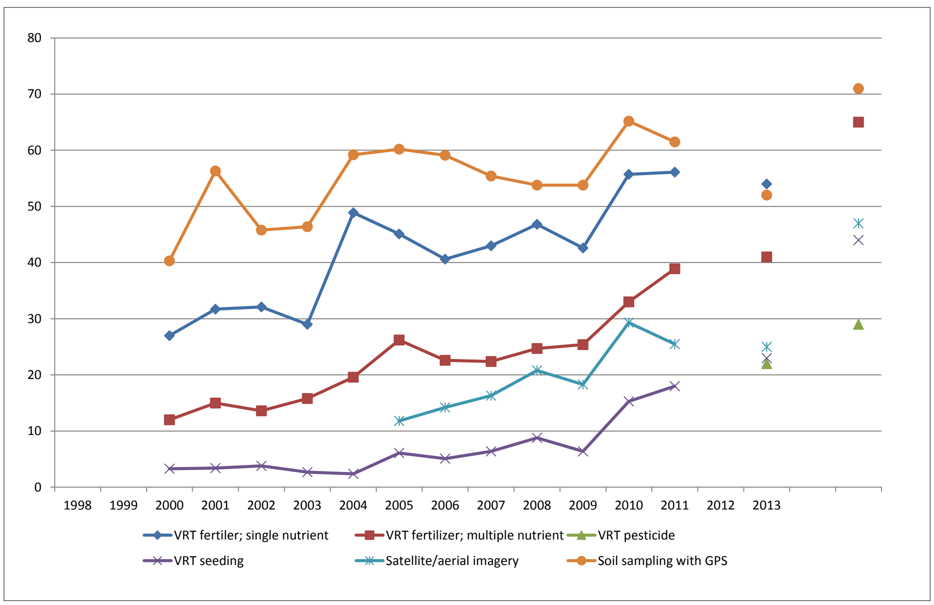

Figure 8. Precision agricultural services offered by input suppliers in the Midwest; 2000-2013

Data on precision agriculture services offered to farmers (Figure 8) tracks the tools used by input suppliers; the most commonly offered service is soil sampling with GPS (7 in 10 dealers). Variable-rate technology (VRT) application of nutrients is commonly offered, with two-thirds offering that service. Other services, VRT application of inputs and satellite/aerial imagery, showed marked increases in recent years, with about 4 in 10 dealers offering those services. This increase may reflect the recent changes in output prices and input costs.

As shown in Figure 9, roughly three-fourths of the input suppliers surveyed believe that their total precision agricultural program is profitable. The percentage of suppliers who feel that variable-rate application of fertilizer is profitable is even greater. Less than 40 percent, however, believe that data analysis for yield monitors contributes to the profitability of their firms.

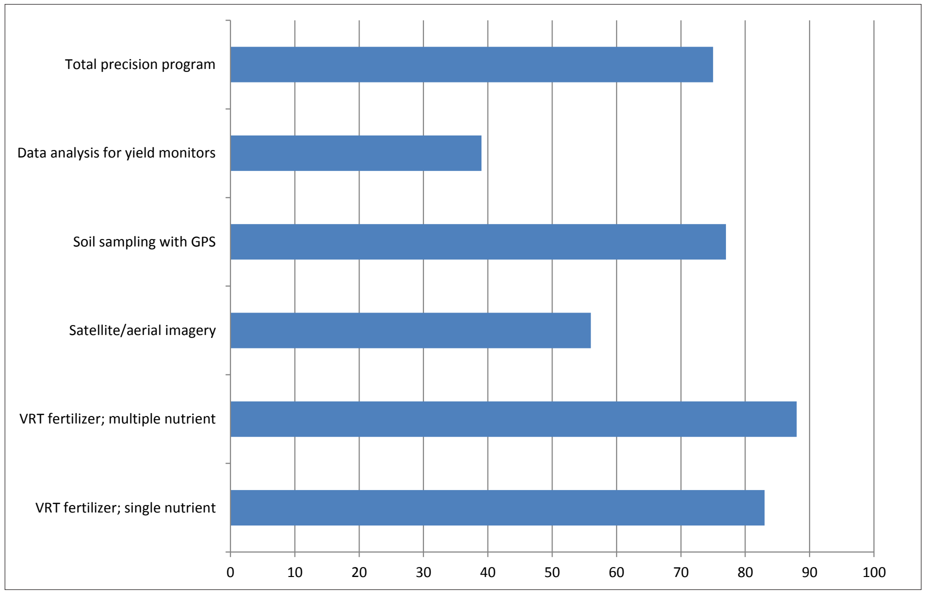

Figure 9. Proportion of input suppliers who feel provision of precision service is profitable

Pendleton[9] provides an integrated view of precision agriculture's role in Midwestern agriculture (Figure 10). This perspective emphasizes the value potential of precision agriculture rather than the technologies involved or specific services provided. With a focus on outcomes, the process extends across creation, capture, analysis and use of data. In so doing, the perspective sets the stage for our later discussions of Big Data.

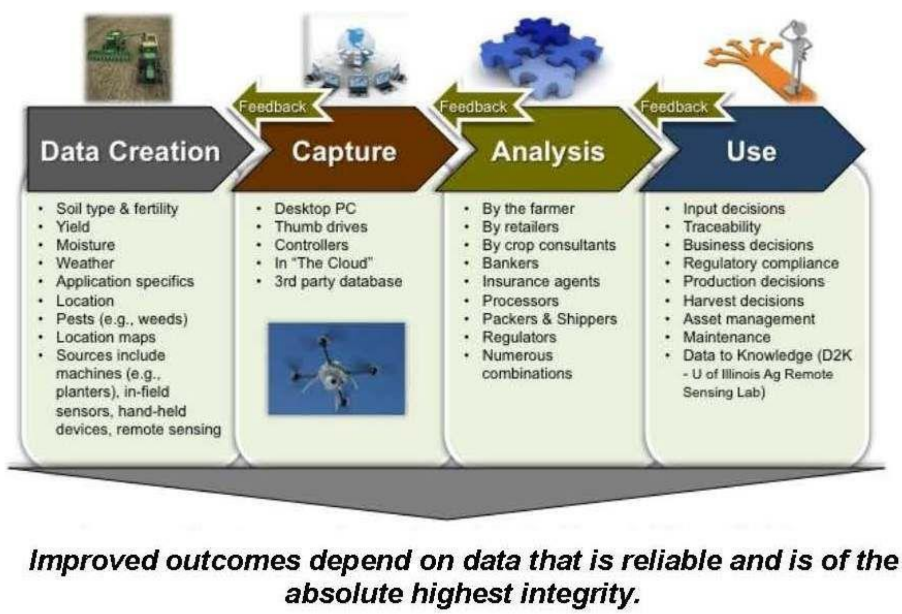

Figure 10. Precision agriculture and value

Emerging Technologies

As noted previously, numerous specific tools and technologies can be included within the broad framework of precision agriculture.

---

[9] Pendleton, K. Precision Ag: A Dynamic Growth Area Redefining Agriculture. Thinking in Context. Summer, 2013. www.Contextnet.com

The data reported in the Purdue/CropLife survey indicates that there is significant use of GPS and VRT-related practices in the Midwest. Not included there are auto-steer capabilities and the automated reporting of engine data as field operations occur. Both are potentially exciting contributors to Big Data initiatives, although less is known about the current rate of adoption.

Figure 11. Irrigated pasture that supports New Zealand's vibrant dairy sector

This section expands our geographic perspective to consider innovations whose current context extends in some cases beyond the Midwest. Each has important potential in the context of a future Big Data-informed agriculture.

Figure 12. Agri-Optics fertilizer application keyed to specific pasture needs

Smart-N Fertilizer Application

New Zealand's dairy sector is a global powerhouse with marketing and collaboration efforts that range far beyond that island nation. Innovation and linkage of new technology to the sector's specific needs have been key features of its long-term success. A key differentiating feature of the sector is its reliance on pasture-based nutrition and significantly less use of grain as a primary source of nutrition.

Abundant pasture throughout the production season is critically important to the economic success of New Zealand dairy producers. Supplemental application of fertilizer during the growing season ensures that an abundant level of grass is available. Of course, additional "contributions" to fertility are made daily as the animals graze. However, because of the nature of the pasture grazing system, that application is done in a rather imprecise fashion. As shown in Figure 11, this results in a non-uniform pattern of grass growth.[10] Traditionally, fertilizer was applied to meet the general needs of each paddock. Doing so meant that more fertilizer was applied than was actually needed.

Figure 12 shows a new technology application, called Smart-N, which was created to address this problem.[11] In general terms, sensors located a few inches in front of the application nozzles

---

[10] Courtesy Craige Mackenzie, Agri-Optics. 2012.
[11] Courtesy Craige Mackenzie, Agri-Optics. 2012.

identify the extent of grass growth and, therefore, the need for supplemental N on the ground directly beneath the sensor. Those needs are then communicated to the nozzle keyed to that sensor. Doing so limits both under- and over-application of fertilizer, thereby reducing costs and fertilizer run-off. Craige and Roz Mackenzie, dairy farmers as well as Smart-N innovators, were recently awarded New Zealand's 2013 Balance Farm Environment Award in recognition of their efforts with Smart-N and other innovations.

Unmanned Aerial Systems (UAS) and Satellite-Based Sensing

To this point, use of GPS and VRT-related tools in Midwest crop farming has been focused on input application at pre-planting, 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 that could change this cost/benefit ratio are emerging.

A recent Wall Street Journal article outlined ongoing efforts to transform UAS technology capabilities originally focused on military purposes to applications supporting production agriculture. "As the spring growing season unfolds, universities are already 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)
- 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 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) 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 ongoing 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. As illustrated in Figure 13, the system employs Synthetic Aperture Radar (SAR) technology, which offers an effective alternative to optical observation, which can be obstructed by clouds, thus putting it at a disadvantage.[12] 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. Figure 14 depicts the RIICE vision, in which satellite-based measurement can be converted to mapping and forecasting tools to provide important information and data-based products for the small-scale rice producers of Southeast Asia.

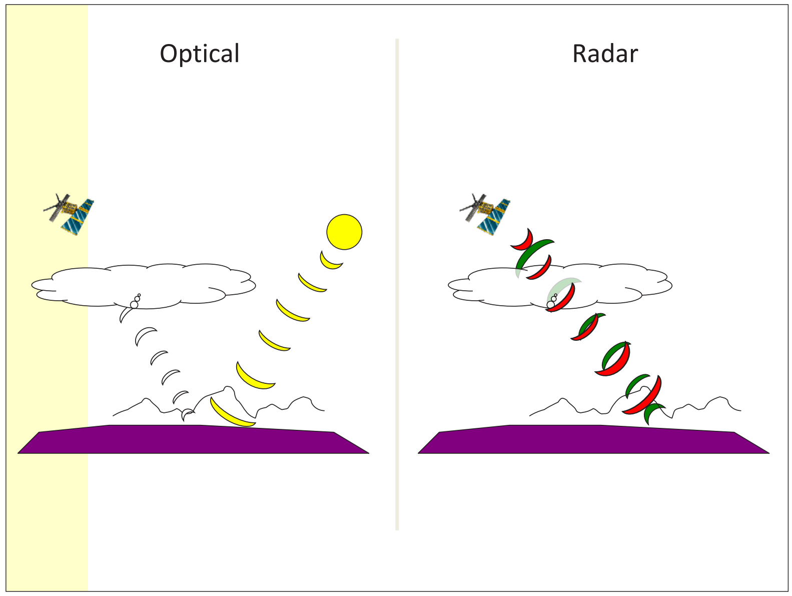

Figure 13. Illustration of superior performance of radar versus optical sensing from satellites

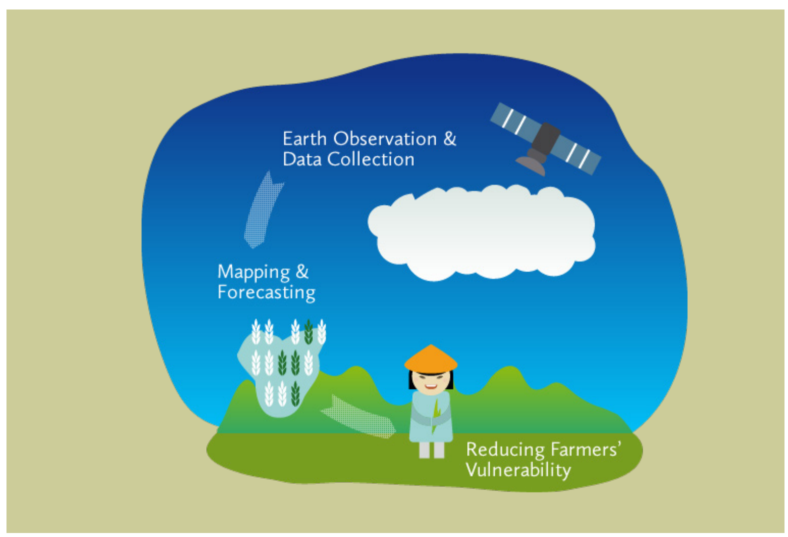

Figure 14. RIICE concept to provide information to reduce vulnerability of rice producers

The Cell Phone as Versatile Sensor

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.[13]

[12] About RIICE. http://www.riice.org/about-riice/

[13] International Telecommunication Union. (2013). Key ICT indicators for developed and developing countries and the world. Retrieved from: http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2013/ITU_Key_2005-2013_ICT_data.xls

Examples of the reach of cell phone use by 2012 include:[14]

- India
  865 million mobile phone subscribers
  #2 Global Market (After China)
- Brazil
  248 million mobile phone subscribers
  #6 Global Market (After China, India, USA, Indonesia and Russia)
- Nigeria
  113 million mobile phone subscribers
  #10 Global Market

Adoption is extensive in urban areas; however, cell phone use is common in rural areas. For example, early in 2001, I had the pleasure of interviewing a group of soybean farmers in Central India. As we discussed their challenges, I asked, "What's the one thing that you'd like to have that would make you more effective as a soybean producer?" The consensus, which was offered immediately, was that they would like to be able to know the price of soybeans on the Chicago Board of Trade. They knew that global price movements affected the local price they were being offered, but had no effective means to track global prices. In 2012, I was again interviewing a group of Indian soybean farmers in the same state. As part of the discussion, I asked, "How do you get price information?" A majority of the farmers held up a cell phone and said, "We get text messages sent to us."

Cell phones, of course, are not created equal. They range from devices that receive phone and text messages to smartphones that effectively operate as small computers. In addition to receiving data and information, the use of cell phones as sensors is gaining 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: when aggregated, they can provide important and timely insights. Figure 15 illustrates this point.[15] 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 15 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 monthly volume of tweets about the price of rice in Indonesia. 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.

[14] 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
[15] Mock, N., N. Morrow, and A. Papendieck. From Complexity to Food Security Decision-Support. Global Food Security. (2013). http://dx.doi.org/10.1016/j.gfs.2012.11.007

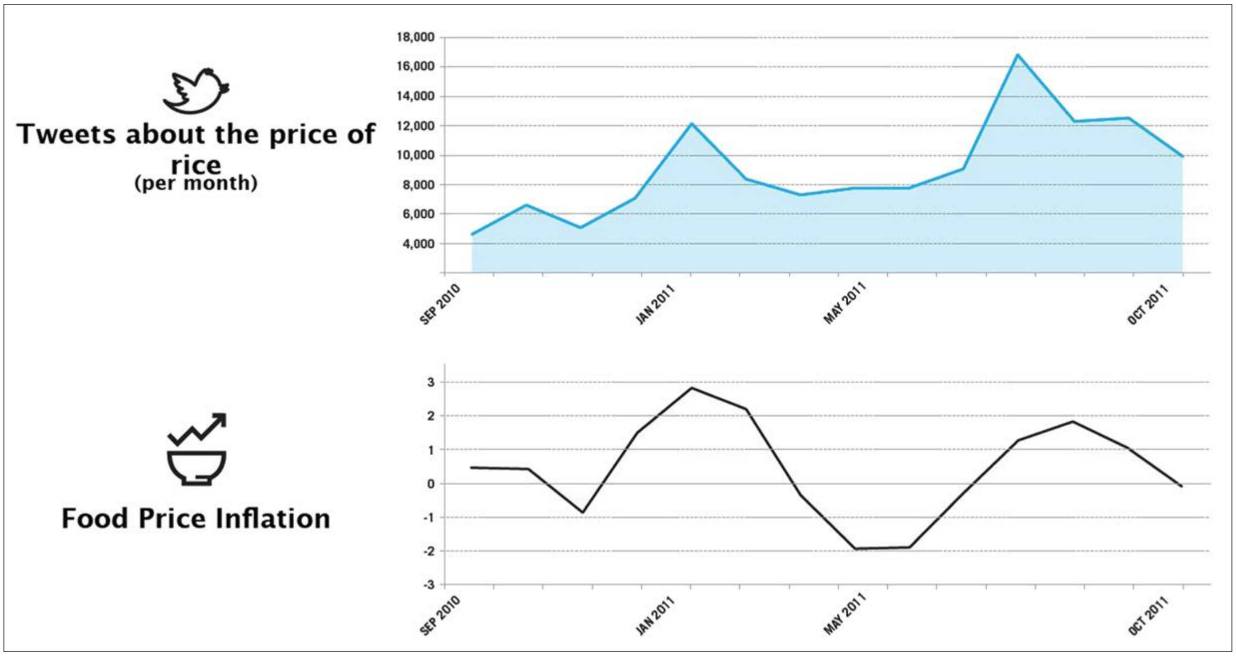

Figure 15. Comparison of estimates of food price inflation and tweets about the price of rice in Indonesia

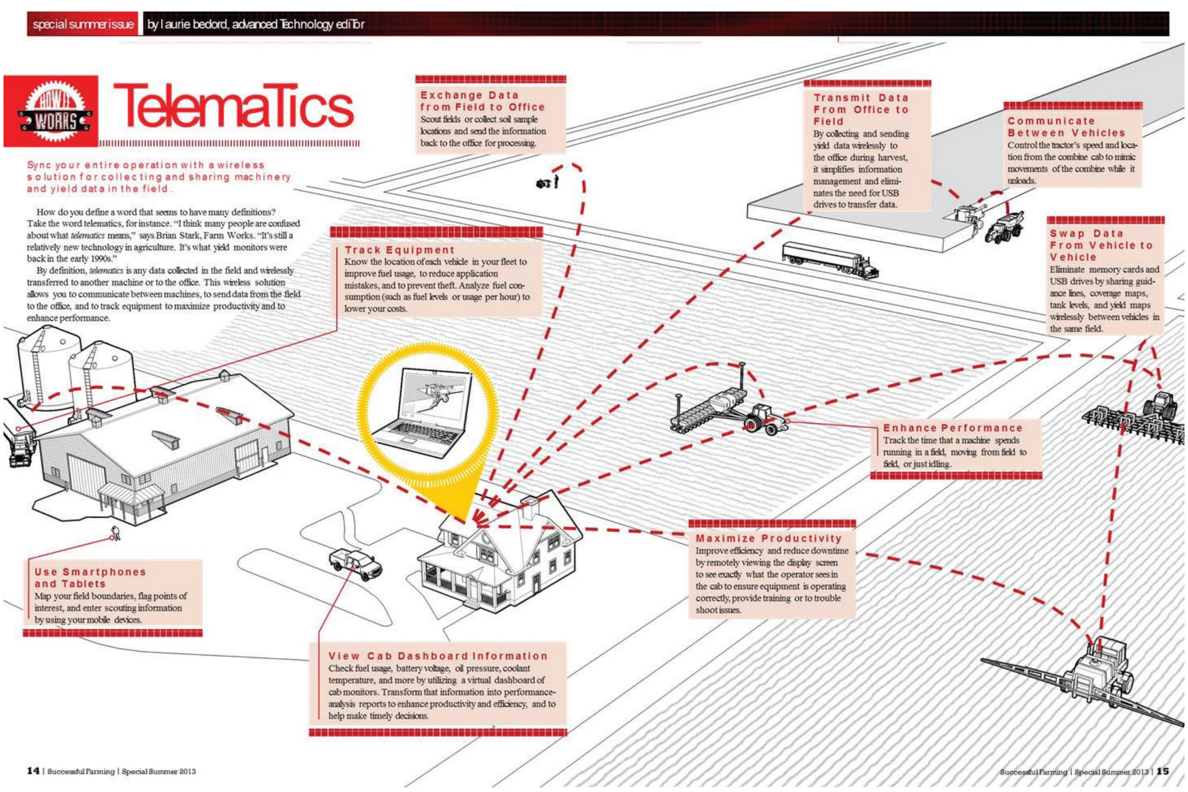

Figure 16. Telematics connects the farm firm

Telematics

Wireless communication isn't limited to conversations between people. Telematics arises from the blending of computers and wireless telecommunications. Figure 16 illustrates a farm-of-the-future where telematic capabilities allow the farm's network to enable rapid and accurate transfer of information among the farm's computers, monitors and sensors.[16] Do you remember when floppy disks where the answer to transferring information from the yield monitor to the home computer? Of course, the most common example of telematics may be the Internet itself, as it depends on a number of connected computer networks across the globe.

Open Data

Significant efforts also are underway to make existing public sources of data more available. U.S. Secretary of Agriculture Tom Vilsack said, "Never before has the world been able to collect so much data on such a wide range of topics — from weather conditions to crop growth to nutrition." He also said, "Data in isolation is not as powerful as data shared."[17] These comments were made at the official launch of the USDA's Food, Agriculture and Rural Virtual Community on the website data.gov. These projects mirror the efforts of several countries worldwide to make public information more readily for available as data as well as interpreted information.

ICT Application and Industry Transformation

Although of great potential importance, the phrase "Big Data" is only the latest buzzword 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 reality of ICT-based advances enhancing business operations and disrupting industries is not just a recent event. Indeed, for managers in the 1990s, the challenge of dealing with and taking advantage of the "knowledge economy" paralleled the attention devoted to Big Data today. At that time, there was considerable noise regarding the advent and potential effects of the knowledge economy. However, several of the key concepts that emerged have proven to be applicable in the long run. Central among these findings is the notion that, while new technology is essential, the increase in learning that results from the technology is the most important element of the process.

Sampler and the Redefinition of Industries

Extensive use of information technology has redefined industries throughout the economy. Although the effects have often been described in some detail, the underlying mechanisms that fuel industry transformation have been more difficult to understand. Sampler[18] provides

[16] *Successful Farming*, Special Summer Issue 2013.
http://www.agriculture.com/uploads/assets/promo/external/pdf/Telematics.pdf
[17] Fatka, J. Open Data Key to Future Ag Innovation. Feedstuffs. May 6, 2013.
[18] Sampler, J.L. Redefining Industry Structure for the Information Age. Strategic Management Journal. 1997.

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.
- 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 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 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, monitors can alert decision makers of the potential for critical problems before these problems occur.

The second transaction characteristic is aggregation potential. Knowing the purchasing habits of one consumer is interesting, but that information provides little economic value. However, being able to accumulate and analyze the purchasing 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 widespread application of the algorithms defined. As detailed by Shapiro and Varian,[19] 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, where we'll also introduce the word "analytics" as the 2013 extension of Sampler's aggregation potential.

---

[19] Shapiro, C. and H.R. Varian. Information Rules. 1999.

Affinity (Loyalty Card) Programs

Early advances in the application of information technology focused on affinity programs for consumers (often using loyalty cards). Here, however, we will consider an agricultural application that failed to succeed because low-cost separability was not met.

The American Airlines "frequent flyer" program (starting in the 1970s) was an early success story. By providing the opportunity to earn benefits, American Airlines created the incentive for its customers to prefer that airline and become repeat customers. Over time, we, as consumers, became used to providing our personal frequent flyer number to the travel agent, then to the reservation clerk and then on the airline's website.

In this context, the transaction is the purchase of an airline ticket. Including a frequent flyer number adds a relatively minor cost. Yet, from another perspective, if the identity of the purchaser is the only information added, there is little gained from use of the frequent flyer number. Even in pre-9/11 days, airlines had to know who was flying on their planes. But use of the frequent flyer number is a low-cost means to tie the transaction to another set of information: the demographic data supplied by the customer when receiving the card. That demographic data can then be linked to other data sets.

But not all loyalty programs have had the success of the American Airlines effort. In the early 1990s, a major agricultural chemical manufacturer initiated an affinity program for farmers who purchased their products. Participating farmers would receive rewards based upon the amount purchased. At that time, agricultural chemical manufacturers had relatively little information about the consumer of their products, the individual farmer. Therefore, an effective loyalty program was attractive as an important source of potential advantage.

In the summer of 1993, I had the opportunity to interview Midwest ag retailers on a number of topics relating to the emerging field of precision farming. Because the loyalty program noted above was relatively new, these discussions often explored the actual means by which transaction information was captured. In this case, the transaction was the purchase of ag chemical products from a specific manufacturer. That transaction occurred between the ag retailer and the farmer customer; therefore, the manufacturer had to rely upon employees of a third party to capture the relevant transaction data. This presented a number of hurdles, including:

- The ag retailers' computer systems were not designed to separately track the relevant data. Therefore, capture of the information for the loyalty program did not occur as part of the financial transaction.

- The transaction occurred at a busy time for the ag retailers' employees. Therefore, the process of recording the loyalty program information was often delayed, sometimes for weeks or months. The delay in recording resulted in errors between the information collected and what actually occurred.

- Because farmers often purchase supplies from more than one retailer, having data from one retailer did not necessarily provide comprehensive customer information. In addition, differences in land rental arrangements, for example, meant that knowing who paid the bill might not provide an accurate linkage to the actual decision maker.

For these reasons, and probably others, that particular loyalty program was not successful. Of course, all innovative efforts don't succeed. The purpose for mentioning this anecdote is to focus on the separability characteristic and the importance of that characteristic for initiatives that relate to production agriculture. Experience shows that not only must it be possible to capture information, but also, the capture of data must be done cheaply, accurately and within normal business operations. Historically, the fragmented nature of agriculture and agribusiness, which encompasses many firms and decision makers, has made it difficult to achieve this characteristic.

Big Data and Its Business Application

Beginning a section titled Big Data seems a lot like jumping into the deep end of the pool. Having grown up on a farm 30 miles from the nearest swimming pool, the deep end isn't a comfortable place for me. So, when I do that, I suspect that there's going to be a lot of floundering around. In this section, we're going to focus on the following three components of the Big Data pool, and hopefully, we will keep our heads above water while doing so:

- 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?

Figure 17's advertisement from IBM illustrates the types of messages that have become common. Though it's exciting and it offers potential gains, it's not necessarily clear what makes Big Data "big" or how to effectively employ it. The examples noted in this section will not be from production agriculture or the food chain. Linking the concepts noted here to potential application for Midwest agriculture will be the focus of the paper's following section.

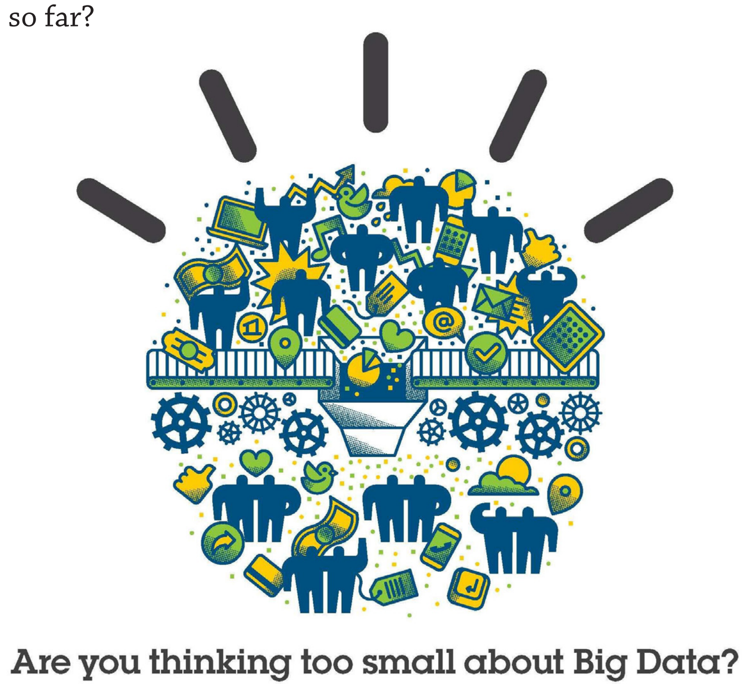

Figure 17. A Big Data advertisement

The reference source of choice for today's college student (Wikipedia) highlights six examples of Big Data use:[20]

- 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 one million customer transactions every hour, which is imported into databases estimated to contain more than 2.5 petabytes (2,560 terabytes) of data — the equivalent of 167 times the information contained in all the books in the U.S. 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 worldwide.
- 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.

Three dimensions are typically used to describe the Big Data phenomenon: volume, velocity and variety. Each dimension presents 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.

The McKinsey Global Institute[21] describes the dimension of volume in Exhibit 2:

> "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 data — i.e., we don't define big data in terms of being larger than a certain number of terabytes (thousands of gigabytes). We assume that, as technology advances over time, the size of datasets that qualify as big data will also increase. Also note that 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).

Exhibit 2: Volume as it relates to Big Data

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

[20] Wikipedia. Big Data. http://en.wikipedia.org/wiki/Big_data data retrieved 6/28/2013.
[21] McKinsey Global Institute, 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 2011.

important. McAfee and Brynjolfsson[22] 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, data refers to numbers meaningfully arranged in rows and columns that can then be summarized in appropriate standardized reports. Usually, these numbers summarize operating and financial performance.

For Big Data, the concept of "what is data" is wildly expanded. We now refer to the "Internet of Things" — the 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 sources of information — of what they are, where they are and where they're headed.

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 the definition of data and the ability of businesses to capture, transmit and store data. What is separable today is greatly different than it was just a decade ago. But, we still have to aggregate the information 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 online illustrates the change of perspective. Tracking 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 online bookseller as well. However, the online bookseller can track and influence the shopping experience of each customer. The online bookseller can track which 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.

---

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

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 allow 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.

Exhibit 3. Monitoring manufacturing processes

Example Applications

We just explored the differing perspectives available to a physical bookstore versus its online competitor, where the insights gained regarding each customer were stacked in favor of the online offering. In Exhibit 3,[23] examples of manufacturing applications are described.

Recently, public attention was fixated on another Big Data application,[24] 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. Exhibit 4 contains excerpts from an article describing the advances in technology that facilitated this Big Data application. Noted there are the software and hardware advances that drive rapid assessment of massive amounts of unstructured data.

---

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

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

Database systems

Traditional databases, usually written in a language known as SQL (pronounced sequel), store data in tables, columns and rows but are limited when it comes to storing strings of words such as those found in an email or text message. They also can't handle pictures or video.

New types of databases that emerged beginning in late 2009, known collectively as NoSQL (for "not only SQL"), such as MongoDB, Cassandra and Simple DB, don't have these limitations, and allow analysts to create queries against all these types of data.

NoSQL databases can make a huge difference to companies analyzing very large data sets, even if they're fairly conventional. For example, analysts at risk consultancy Verisk Analytics Inc. are "constantly running different models and analytics" against billions of customer records in order to help identify fraudulent insurance claims.

Perry Rotella, vice president and chief information officer at Verisk, says using a traditional DB2 database from International Business Machines Corp., "would be a six-hour job" that had to run overnight. Analysts would pore over the results and generate new queries that would again have to run overnight. He said it took weeks every time analysts needed to create a new statistical model. The company recently changed to a NoSQL database that allows analysts to run the same types of queries in 30 seconds.

"So all of a sudden your model-building becomes iterative in real-time instead of over days. [Using NoSQL], you can run analytics on your data multiple times a day, and it compresses your ability to get results from weeks into days. It's extremely powerful," he said.

Machine learning

Traditional analysis requires analysts to have enough understanding of the data to form a hypothesis and then create complex queries to run against the database. Recently developed programs known as machine learning and natural language processing rely on the computer programs themselves to find patterns and even elucidate the meaning of ambiguous words based on context. "You can turn a machine-learning program loose on a lot of data and you can see what they are able to be predictive of," said Mr. Davenport, a visiting professor at Harvard Business School. With natural language processing, "you could figure out whether a term like 'bomb' is being used to describe a Broadway play versus something a terrorist would use," he said.

Machine learning, also known as cognitive analytics, allows queries to continually "tune themselves," Gartner Inc. analyst Douglas Laney explains. For example, retailers use this technology to automatically update pricing algorithms in real time as new information, such as weather, time of day and even information gleaned from video of customers browsing in their stores become available. "It used to take more than a day to update pricing, but these retailers can reprice every hour and use trending information to do real-time product pricing," says Mr. Laney. "I'm not sure they could do that even a year ago," he said.

Hadoop

Until recently, complex computer programs needed to run on expensive hardware, such as enormous mainframe computers. Today, an open-source software framework called Hadoop — developed at Yahoo Inc. with contributions from technology developed by Google Inc. and named after a child's toy elephant — allows queries to be split up by the program, with different analytic tasks distributed among scads of inexpensive servers, each of which solves a part of the puzzle, before reassembling the queries when the work is completed. "It's really cheap and really fast," said Mr. Davenport.

The ability to distribute complex queries to a large number of inexpensive computers helps people get very quick responses to complicated questions with a large number of variables. For example, online automotive market (Edmunds.com Inc.) can help auto dealers predict how long a given car will remain on their lots by comparing car makes, models and trim against the number of days inventory cars at that price point averaged on a lot in a given dealer's region. The predictions help minimize the number days a car remains unsold — "one of the most important sales metrics for dealers," said Philip Potloff, Edmunds.com's chief information officer.

Exhibit 4. Software and hardware innovations fueling Big Data use

Managerial Lessons

While aggressive new uses for ICT have been a constant for the last two decades, the pace and nature of the change itself has been evolving. Further, it often is difficult to separate the contribution of ICT use from the investment and operating decisions necessary to achieve economic gain. Without changes in production, distribution and marketing, an enhanced information capability often won't generate value.

A recent analysis conducted by the Economist Intelligence Unit indicates that business leaders feel there is a strong link between the use of Big Data and economic performance.[25] 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
- 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[26] 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. They include the ability to:

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

Big Data and Midwest Agriculture: Vision and/or Hallucination?

Big Data applications are employed throughout the economy and society. The technologies are exciting, involve analysis of mind-numbing amounts of data and require fundamental rethinking as to what constitutes data. 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

---

[25] Economist Intelligence Unit. Big Data: Lessons from the Leaders. 2012.
[26] Barton, Dominic and David Court. Making Advanced Analytics Work for You. Harvard Business Review. October, 2012.

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 Midwest agriculture clearly 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.[27] The first question was, "How will IT change the basis of competition in your industry?" Exploring, but not answering, that question is the purpose of this section. While we expect Big Data applications to be used within Midwest agriculture, it doesn't mean their use will change the basis of competition. Indeed, the question in this section's title will hopefully reminds us that there's often a thin line between vision and hallucination.

The following discussion briefly reviews the key distinctive aspects of Big Data. Then, we'll explore two innovations that employ Big Data to develop potential services for agriculture. Finally, we'll discuss the issue of industry transformation 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. While media descriptions of Big Data enjoy focusing on unfamiliar terms describing the quantity of data, McKinsey's more meaningful description focused on "data sets whose size is beyond the ability of typical database software tools." Velocity refers to the capability to have information that can inform decision making in "real or near-real time." 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 webpages 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 satellites
- Questions

[27] P. Wilmot. The do-or-die questions boards should ask about technology. McKinsey & Co. 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

Analytics, Algorithms, Proxies and DIRT!

Massive increases in computational power enable analysts today to do more than crunch financial information. As we've seen, data can be almost anything that can be assessed. Some key terms are:

Analytics — "making sense" of massive amounts of highly variable data types
Algorithms — a set of instructions that starts from an initial state and initial input and proceeds through a finite number of well-defined successive states, eventually producing output.
Proxy — a substitute for another

One aspect of the power of Big Data is the capability to employ extensive computation capabilities to explore very large data sets. The goal of doing so is to discover insights and relationships that can be used to anticipate future behavior of complex phenomenon.

Let's speculate about an instance where analytics might discover agriculturally important information, if we had access to lots of data. First a bit of algebra, in the form of equation 1:

(1) $Z = c + bX$

But let's say we can't get good data on the factor shown as c in (1) because it's too expensive to measure. However, if we happen to know values for Z, b and X we can determine c, as in equation (2):

(2) $c = Z - bX$

In Midwest corn production, we know that yields are determined by factors such as soil capabilities, inputs applied, tillage practices, planting date and weather events during the growing season (This is intentionally a simplified list). However, farmers and managers have difficulty capturing precise data on soil capabilities at the same level of detail as they apply inputs and measure yields.

What if an analyst had massive amounts of GPS referenced data on yields, input applications, and weekly crop growth during the growing season? (Here soil capabilities are the equivalent of c in Equation (2)). It should be possible to produce maps of soil capabilities at the same level of GPS detail. And, existing soil type information also could be incorporated into the algorithms, if doing so aided the analysis.

In this speculative example, crop growth could be used as a proxy for weather events. These variables are important when the proxy has a known correlation with the variable of interest. Earlier, we saw an example where cell phone usage near the parking lots of large retail outlets was used as a proxy for the volume of shopping activity on Black Friday.

What if we employed Big Data techniques to develop precise soil capability relationships from, say, 100,000 acres across the Midwest? Could those relationships provide valuable estimates (proxy values) for use all across the Midwest?

Exhibit 5. Measuring what we can't measure

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, is equally important (see Exhibit 5). Our history conditions us to frame analysis as a process where data is assessed to provide an answer. Powerful analytical tools today can search unstructured data with the goal of "identifying the questions" embedded in the data. Electronic communication, which transmits data and and disseminates information, is necessary to achieve velocity.

Two Big Data Innovations for Agriculture

Figure 18 displays a rendition of the Lettuce Bot, a technology being developed to identify and then eliminate weeds in the field.[28] Developed at a California startup, 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 18. 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 19 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 19 cites numerous potential benefits of using Deep Thunder-based services. These include direct linkage to precision agricultural practices, increased yields, reductions of post-harvest loss and consumer benefits of lower price, improved quality and presumably lesser environmental impact. A key factor noted is the potential for more effective water use, a critical global concern.

---

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

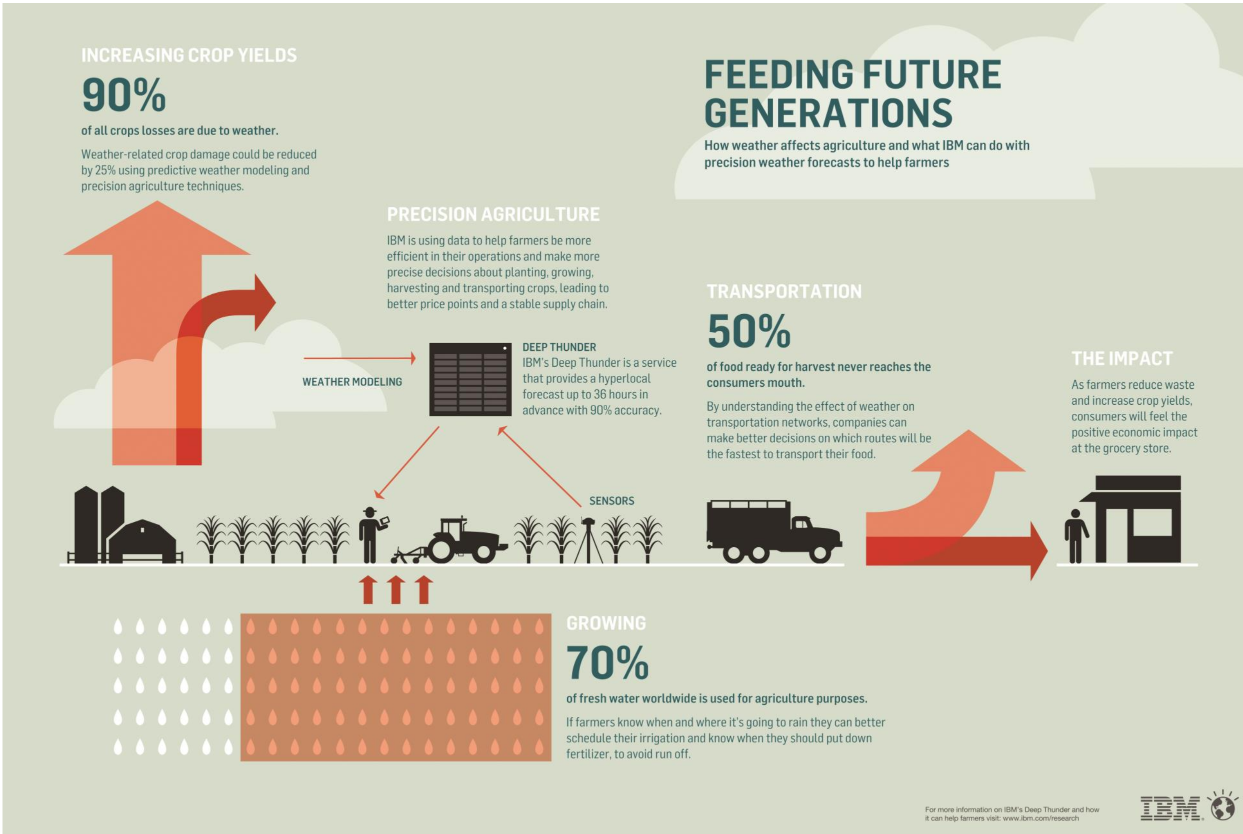

Figure 19. Depiction of use of Big Data technology to provide hyperlocal weather forecasts to support agriculture

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 independent of agribusiness operations. Individual managers can assess the value of the offered service and then adopt the service or not. The Big Data basis of the service is essentially irrelevant to the manager making the purchasing decision. Such innovations will affect competition in specific markets, but are not likely to alter the basis of industry competition. Alternatively, three examples of altered competition include:

- Walmart 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 20 provides a highly useful illustration of data integration and analysis as an attractive vision for production agriculture.[29] In some respects, it is similar to the CCNetAg precision farming graphic shown earlier. Both graphics emphasize data collection relative to field operations and to harvest. Figure 20 has one important point of difference: its explicit emphasis on analytics and integration versus the implied visual nature of the prior precision farming graphic. Indeed, Figure 20 is much more explicit about the need to integrate across many production sites rather than to expect to make substantial advances based upon precise data from a single field.

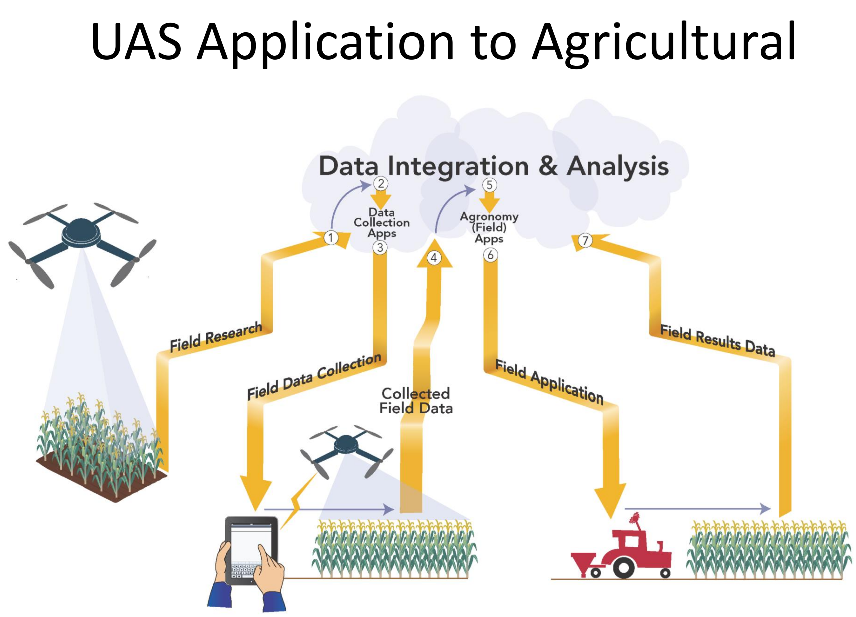

Figure 20. Representation of data integration and analysis for production agriculture

Figure 20, although useful, masks two key factors that will have to be addressed if Big Data applications are to alter the basis of competition in Midwestern agriculture. One factor relates to the word integration, particularly the process of integrating across numerous sources of data in which the control of those data doesn't exist within a single entity. The second factor relates to the word analysis. Figure 20 suggests that the necessary data can be obtained from field operations, just as counting the

[29] Courtesy Riverside Research, Champaign, Ill.

turns of a screw can determine if manufacturing processes are within bounds. Field data, while essential, is likely insufficient 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?

Midwestern commercial farms continue to grow in size, but the individual farm likely won't have sufficient scale and variety of operations to create significant new information to optimize operations. Indeed, what the individual farm 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 farm's operations. VRT application of inputs often is done by agricultural retailers. Although the producer typically does the planting, 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 20 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. Midwest 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 upon 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 ongoing 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 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. [30]

Knowing, at increasing levels of precision, "what" happened in the field or in the animal facilities does have value. However, not also knowing "why" is concerning 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.

From the perspective of pure technological capability, these aggregation challenges are not insurmountable. However, the associated economic and organizational issues likely will be significantly more difficult to overcome:

- 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.

- 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 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 record systems operate to achieve similar goals to those noted here. Also, it should be noted that efforts are underway to develop technologies and protocols where major non-ag manufacturers can "share" operational data to allow learning to occur, but not compromise intellectual property or competitive factors.

---

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

Wrapping It Up

It seems that one can't go through a typical workday without seeing some mention of Big Data, its application and its potential to have unprecedented impact. Agriculture, globally and in the Midwest, 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.

The purpose of this case study is to explore the potential for Big Data in agriculture, especially in the Midwest (see Exhibit 6). ICT-enabled innovation in Midwest agriculture is not new. Starting in the early 1990s, initiatives linked to precision agriculture have been adopted throughout the Midwest. Facilitated by GPS technology, practices are available to more effectively plan, apply and monitor input application. These variable-rate techniques are employed on a widespread basis. GPS-based tools to monitor within-field yield variability at harvest are commonly available. Increasingly, auto-steer capabilities are enhancing effectiveness of within field operations. Continued adoption of these tools indicates their effectiveness, particularly in advancing operational efficiency.

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 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 business issues and opportunities, not technological capabilities, which determine success. That counsel seems appropriate for Midwest agriculture 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 disconnected from agricultural operations to provide enhanced product and service offerings to the sector. The choice whether to adopt these offerings will be 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. While there will be technology impediments, key economic and organizational issues likely will be more challenging. Data

The Third Tell

Let's return to the phrase, "Surprise is inevitable, Being unprepared is not." Below I speculate about some key concepts to consider to be better prepared for Big Data application in Midwest agriculture.

Advances in information and communication technologies (ICT) have had massive effects in society. The cell phone, the Internet and the set of technologies we know as precision agriculture have already had significant impact in agriculture. The adoption of precision agriculture in Midwest farming has grown significantly, particularly in recent years. As farm size continues to increase the use of precision agriculture-based practices likely will expand. The widespread use of these practices will enhance the potential for Big Data application.

The three key dimensions of big data are volume, velocity and variety. Of these three, variety has the potential to be the source of significant surprise to managers because what is now data is very different than we're used to. Conversations that mention the price of rice (expressed as tweets) now can indicate price movements of that essential food in developing nations. Analytics, the ability to make sense of massive amounts of highly variable types of data, is a key source of the potential power of Big Data. The potential surprise is that a competitor firm may be making more effective decisions based upon data — that others in the industry don't realize is even being measured.

Let's consider three paths by which Big Data might enter agribusiness.

- We've reviewed a number of applications in manufacturing and logistics. As providers become more proficient in applying these tools, the applications will decrease. As these learning effects occur, procedures now only used to produce rockets could be attractive in many manufacturing settings. Consumer marketing is a "hot" venue for linking social media and Big Data. Large food product manufacturers and retailers are striving to employ these techniques to improve their customer offerings.
- Earlier we looked at two product development efforts employing Big Data where the service to be offered didn't rely upon real-time data from the agribusiness operations. We can expect many more such product/service offerings in agribusiness.
- In numerous sectors, ICT-based advances have altered the industry's basis of competition. One means by which Big Data could alter the basis of agribusiness competition would be to foster new linkages across traditional boundaries in the sector. This could involve multiple farming units as well as suppliers and customers of the farming operation. For example, after harvest a major portion of the corn crop resides within distributed storage facilities across the Midwest. At planting, fertilizer inputs are poised for use in another distributed set of facilities. Today, the process by which those goods are brought to market is done in a highly efficient manner, based, however, upon the decisions of thousands of decision makers considering only their independent economic needs.

    - What value, if any, could be created by employing Big Data tools and GPS-based location information to optimize performance among these many firms?
    - If value were created, what organizational structure and business model would facilitate effective initiation of such a system?

Economics will matter. While advances in technology will continue to amaze, economics will have material effect in determining which initiatives are visionary and which turn out to be hallucinations. Key concepts are:

- Low cost data acquisition invariably is a key.
- Analytics has transformed the process of creating potential value through aggregation.

For initiatives that require linkages across traditional agribusiness boundaries, innovations in organizational collaboration and use of our Small Data knowledge base likely will be essential in numerous instances. Taking advantage of such collaboration may require the most preparation.

Exhibit 6. Tell them what you've told them

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 need to be addressed. For many Big Data applications, valuable business insights can be gained from the capture and analysis of operational data only. However, the biological 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 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 Midwest agriculture but that success will be driven as much by business and managerial factors as it will by technology.