International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

Big Data Characteristics

Steven Sonka[ⓘ]

Emeritus Chaired Professor of Agricultural Strategy, University of Illinois, 326 Mumford Hall, 1301 W. Gregory Drive, Urbana, IL 61801, USA

Often, *big data* is referred to as a singular entity. It is not! In reality, big data is much more a capability than a thing. It is the capability to extract information and insights where previously it was economically, if not technically, possible to do so. Advances across several technologies are fueling the growing big data capability. These include, but are not limited to computation, data storage, communications, and sensing. The growing ability of analysts and managers to exploit the information provided by the big data capability is equally important.

Only recently have numerous attempts been made to define big data. For example:

- The phrase "big data" refers to large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, e-mail, video, click streams, and/or all other digital sources available today and in the future (The National Science Foundation 2012).

- Big data shall mean the datasets that could not be perceived, acquired, managed and processed by traditional IT and software/hardware tools within a tolerable time (Chen et al. 2014).

- Big data is where the data volume, acquisition velocity, or data representation [variety] limits the ability to perform effective analysis using traditional relational approaches or requires the use of significant horizontal scaling for efficient processing (Cooper and Mell 2012).

- Big data is high-volume, -velocity, and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making (Gartner IT Glossary 2012).

[ⓘ]Corresponding author: Tel: + 1 217.333.1810
Email: S.S. Sonka: ssonka@illinois.edu

The goal of this paper is to move beyond those definitions to explore the characteristics of big data which have particular relevance in fostering the creation of value in agriculture.

Dimensions of Big Data

Three dimensions (Figure 1) often are employed to describe the big data phenomenon: *volume, velocity, and variety* (Manyika et al. 2011). Each dimension presents both challenges for data management and opportunities to advance agri-business decision-making. These three dimensions focus on the nature of data. However, just having data isn't sufficient. Analytics is the hidden, "secret sauce" of big data. Analytics (discussed later), refers to the increasingly sophisticated means by which useful insights can be fashioned from available data.

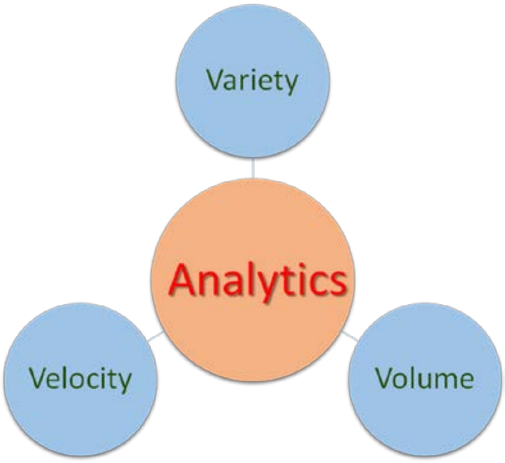


Figure 1. Dimensions of Big Data

Volume: According to IBM (2012) 90% of the data in the world today has been created in the last two years alone. In recent years, statements similar to IBM's observation and its emphasis on volume of data have become increasingly more common.

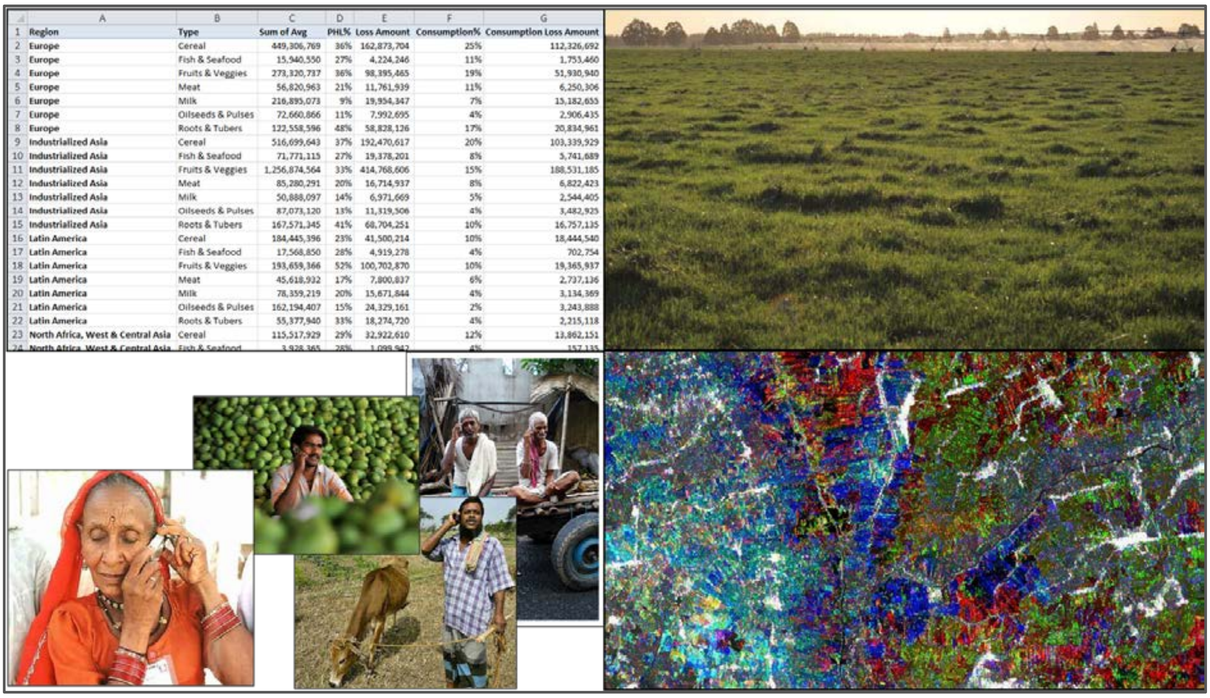
The volume dimension of big data is not defined in specific quantitative terms. Rather, big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze. This definition is intentionally subjective; with no single standard of how big a dataset needs to be to be considered big—and that standard can vary between industries and applications.

Velocity: The velocity dimension refers to the capability of understanding and responding to events *as they occur*. Sometimes it's not enough just to know what's happened; rather we want to know what's happening. For example, applications like Google Maps provide real-time traffic information at our fingertips. Google Maps provides live traffic information by analyzing the speed of phones using the Google Maps app on the road (Barth 2009). Based on the changing traffic status and extensive analysis of factors that affect congestion, Google Maps can suggest alternative routes in real-time to ensure a faster and smoother drive.

Variety: As a dimension of big data, variety may be the most novel and intriguing. For many of us, data refers to numbers meaningfully arranged in rows and columns. For big data, the reality of "what is data" is wildly expanding. For example, the movement of your eyes as you read this text could be captured and employed as data.

Suddenly (at least in agricultural measurement terms), the "what is data" question—the variety dimension of big data—has new answers. Figure 2 provides a visual illustration of the change. In its upper left hand corner, we see data as we are used to it – rows and columns of nicely organized numbers. The picture in the upper right hand corner is of a pasture in New Zealand. Pastures are the primary source of nutrition for dairy cows in that country and supplemental fertilization is a necessary economic practice. The uneven pattern of the forage in that field is measured by a sensor on the fertilizer spreader to regulate how much fertilizer is applied—as the spreader goes across the field. In this situation, uneven forage growth is now data. (This also is an example of velocity where the measurement activity is directly linked to action based upon the measurement.)

The lower left hand corner of Figure 2 shows the most versatile sensor in the world – individuals using their cell phone. Particularly for agriculture in developing nations, the cell phone is a phenomenal source of potential change—because of both the information sent to those individuals and information they now can provide. And as illustrated in the lower right hand quadrant of Figure 2, satellite imagery can measure temporal changes in reflectivity of plants to provide estimates of growth (RIICE 2013). The picture is focused on rice production in Asia.



	Region	Type	Sum of Avg	PHL%	Loss Amount	Consumption%	Consumption Loss Amount
2	Europe	Cereal	449,306,769	36%	162,873,704	25%	112,326,692
3	Europe	Fish & Seafood	15,940,550	27%	4,224,246	11%	1,753,460
4	Europe	Fruits & Veggies	273,320,737	36%	98,395,465	19%	51,930,940
5	Europe	Meat	56,820,963	21%	11,761,939	11%	6,250,306
6	Europe	Milk	216,895,073	9%	19,954,347	7%	13,182,655
7	Europe	Oilseeds & Pulses	72,660,866	11%	7,992,695	4%	2,906,435
8	Europe	Roots & Tubers	122,558,596	48%	58,828,126	17%	20,834,961
9	Industrialized Asia	Cereal	516,699,643	37%	192,470,617	20%	103,339,929
10	Industrialized Asia	Fish & Seafood	71,771,115	27%	19,378,201	8%	5,741,689
11	Industrialized Asia	Fruits & Veggies	1,256,874,564	33%	414,768,606	15%	188,531,185
12	Industrialized Asia	Meat	85,280,291	20%	16,714,937	8%	6,822,423
13	Industrialized Asia	Milk	50,888,097	14%	6,971,669	5%	2,544,405
14	Industrialized Asia	Oilseeds & Pulses	87,073,120	13%	11,319,506	4%	3,482,925
15	Industrialized Asia	Roots & Tubers	167,571,345	41%	68,704,251	10%	16,757,135
16	Latin America	Cereal	184,445,396	23%	41,500,214	10%	18,444,540
17	Latin America	Fish & Seafood	17,568,850	28%	4,919,278	4%	702,754
18	Latin America	Fruits & Veggies	193,659,366	52%	100,702,870	10%	19,365,937
19	Latin America	Meat	45,618,932	17%	7,800,837	6%	2,737,136
20	Latin America	Milk	78,359,219	20%	15,671,844	4%	3,134,369
21	Latin America	Oilseeds & Pulses	162,194,407	15%	24,329,161	2%	3,243,888
22	Latin America	Roots & Tubers	55,377,940	33%	18,274,720	4%	2,215,118
23	North Africa, West & Central Asia	Cereal	115,517,929	29%	32,922,610	12%	13,862,151
24	North Africa, West & Central Asia	Fish & Seafood	3,928,365	28%	1,099,942	4%	157,135

Figure. 2. A few sources of data (Sonka and Cheng 2015).

While satellite imagery is one source of remotely sensed data, recent years have seen a pronounced increase in the capabilities and interest in Unmanned Aerial Systems (UAS) as a source of data for agriculture. There are numerous ongoing efforts to transform UAS technology originally focused on military purposes to applications supporting production agriculture. Further, the potential for proliferation of mini-satellites suggests that remotely captured information may become increasing cost effective for use by agricultural decision makers.

Analytics: Access to lots of data, generated from diverse sources with minimal lag times, sounds attractive. Managers, however, quickly will ask, what do I do with all this stuff? Without similar advances in analytic capabilities, just acquiring more data is unlikely to have significant impact within agriculture. While volume, velocity, and variety are necessary, *analytics* is what allows for fusion across data sources and for new knowledge to be created.

Analytics and its related, more recent term—data science, are key factors by which big data capabilities can contribute to improved performance in the agricultural sector. The differentiating features of big data analytics are 1) inclusion of unstructured and structured data types in combination with 2) extremely large data sets. Data science refers to the study of the generalizable extraction of knowledge from data (Dhar 2013). Tools based upon data science are being developed for implementation in agribusiness, although these efforts are in very early stages.

The concept of analytics is maturing and its uses refined (Davenport 2013; Watson 2013). Analytic efforts can be categorized into one of three types:

- Descriptive efforts focused on documenting what has occurred;
- Predictive efforts exploring what will occur, and;
- Prescriptive efforts identifying what should occur (given the optimization algorithms employed).

In agriculture, as in most fields, descriptive efforts have been most common and even those are relatively infrequent. Within production agriculture, knowing what has occurred—even if very accurately and precisely—does not necessarily provide useful insights as to what should be done in the future.

Production agriculture is complex, where biology, weather, and human actions interact. Science-based methods have been employed to discern why crop and livestock production occurs in the manner in which they do. Indeed, relative to the big data topic, it might be useful to consider these methods as the *small data* process.

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

In agriculture, therefore, knowledge from science will need to be effectively integrated within efforts to accomplish the goals of predictive and prescriptive analytics. Even with this additional complication, the potential of tools based upon emerging data science capabilities offers significant promise to more effectively optimize operations and create value within the agricultural sector.

References

Barth, D. 2009. The bright side of sitting in traffic: Crowdsourcing road congestion data. Google's Official Blog: https://googleblog.blogspot.com/2009/08/bright-side-of-sitting-in-traffic.html

Chen, M., Mao, S., and Y. Liu. 2014. Big data: A survey. *Mobile Networks and Applications* 19(2): 171–209.

Cooper, M., and P. Mell. 2012. Tackling Big Data. National Institute of Standards and Technology. http://csrc.nist.gov/groups/SMA/forum/documents/june2012presentations/fcsm_june2012_cooper_mell.pdf [accessed February 2, 2016].

Davenport, T. H. 2013. Analytics 3.0. *Harvard Business Review* 91(12): 64–72.

Dhar, V. 2013. Data Science and Prediction. *Communications of the ACM* 56 (12): 64–73.

Gartner IT Glossary. 2012. *Big Data*. http://www.gartner.com/it-glossary/big-data.

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

IBM. 2012. *What is big data?* https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Byers. 2011. *Big Data: The next frontier for innovation, competition, and productivity*. McKinsey & Company website: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation. [Accessed February 2, 2016].

National Science Foundation. 2012. Core techniques and technologies for advancing big data science & engineering. National Academies Press. http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.htm. [Accessed February 2, 2016].

Remote sensing-based information and insurance for crops in emerging economics [RIICE]. 2013. About RIICE. International Rice Research Institute. http://www. riice.org/about-riice/ [Accessed February 2, 2016].

Sonka, S., and Y.-T. Cheng. 2015. Big Data: More Than a Lot of Numbers! *farmdoc daily* Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign. (5):201 http://farmdocdaily.illinois.edu/2015/10/big-data-more-than-a-lot-of-numbers.html.

Watson, H. J. 2013. The Business Case for Analytics. *BizEd* 12 (3): 49–54.

