

Can Agricultural or Farm Data Be Considered Big Data?

Terry Griffin (twgriffin@ksu.edu)

Kansas State University Department of Agricultural Economics - September 2017

I am often asked if farm data are truly 'big data'. It's a valid question, especially since the nature of agriculture does not fit neatly into existing categories for most any discussion. So, does farm data qualify as 'big data'? In short, yes; here's why. The consensus among agricultural data experts is that farm data are 'big' once data from individual farms are pooled together with farm data from other farms into a community. Even the largest American farms would still be considered 'small data' when in isolation (see forthcoming blog post for more detail). Applying the criteria suggested by [Victor Mayer-Schönberger and Kenneth Cukier](#) to the farm data community, the three V's of big data (Volume, Velocity, Variety, and Veracity) hold (OK, maybe four V's; [see Gartner for more details](#)). Here, I build upon these discussions regarding how the Vs of big data apply to agriculture.

Volume: a lot of farm data exists on office computers, cloud services, and more being collected daily. It has been estimated that 10 MB of data per acre may be collected from farm equipment for planting, spraying, and harvesting; and when only looking at the 90 million acres of corn in the United States 900 TB (that is terabytes; 1 TB is 1,000 GB) could be collected in just one season. [Dr. Scott Shearer](#) puts the volume of farm data into perspective by reporting the annual storage requirement not as TB per season but as 0.5 KB per corn plant. In 2017, the Ohio State University Precision Agriculture team are conducting the "Terra Byte" project; recording how much data has been collected. When considering multiple years of farm data, especially for all commodity crops not just corn, the community dataset is much too big to move around via broadband connection speeds or even with external hard drives; therefore the analytics must be moved to the data. In this agricultural example, the first V (Volume) has been met.

Velocity: meaning changing quickly. Looking only at [as-planted data collected from planters](#) via telematics, 5.5 MB of data on location, speed, cultivar, and other geo-spatial and meta-data are collected for each acre. During planting seasons, the size of the aggregated farm data community becomes much larger every day. Although agricultural operations are seasonal, it should be recognized that even for commodity crops like corn, cotton, soybean, rice, and wheat that peak planting times differ for each such that as-planted data are collected during several months of the year rather than all at once. In addition to planting, other field operations such as tillage, spray applications, and harvest occur at other times during the season; each operation adding to the community of data. Again, this agricultural example meets the 2nd V (Velocity) of big data.

Variety: the spectrum of data sources. In the previous examples of file sizes, each of those in-field operations may be accomplished with different brands of equipment and thus different proprietary file formats. In addition, third party aftermarket telematics may have been added. In addition to the near-automated data collection and transfer from machine-based sensors and telematics, farm data include



data collected via manual methods such as soil sampling that are analyzed by a wide variety of laboratories (see the [Modus Project](#) for overview of efforts to standardize data from labs). Other farm data may be collected and stored in a wide variety of unstructured formats. The variety of farm data sources rivals that of most any other industry and also leads to increased veracity (the 4th V); and therefore clearly meeting the 3rd V of big data.

Veracity: Data are messy in general, and farm data is no exception. In practice, farm data may be messier than most data scientists despite several attempts by the agricultural industry to remove barriers resulting from cross platform use. Data quality has been a contentious topic in precision agriculture for decades; especially regarding raw yield monitor data and other farm data collected by on-the-go sensors. A part of the debate on the veracity of yield data involves whether the farmer or combine operator properly calibrates the yield monitor. Therefore, both sensors and human error influence farm data quality. Other potential sources impacting farm data quality involve manual entry such as cultivar or other input products when making applications. The rate of a liquid can be automatically measured and recorded for gallons per acre being applied, but less automation exists for the actual product being applied. For the 4th V (Veracity) agriculture not only meets the criteria but probably defines messiness better than other industries using big data.

These aforementioned problems of farm data support the notion that agricultural data are truly big when it comes to data; therefore providing opportunities for agriculturists poised to make the most of these technologies. The four Vs of big data readily describe the challenges that agricultural data scientists are facing. Regardless of how big data is defined or how agriculturalists make use of farm data, substantial opportunity remains for proving the value of big data for farmers, their advisors, and society as a whole. These topics will be addressed in the forthcoming articles posted here.

References

Mayer-Schönberger, Viktor and Cukier, Kenneth. 2014. Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt Publishing Company, New York, NY. 257 pp.
<http://www.amazon.com/Big-Data-Revolution-Transform-Think/dp/0544227751>

Modus Project. <https://modusproject.atlassian.net/wiki>

Shearer, S.A. 2014. Big Data: The Future of Precision Agriculture. Presented at The InfoAg Conference. St. Louis, MO, June. http://past.infoag.org/abstract_papers/papers/paper_233.pdf

For more information about this publication and others, visit AgManager.info.
K-State Agricultural Economics | 342 Waters Hall, Manhattan, KS 66506-4011 | 785.532.1504
www.ageconomics.k-state.edu
Copyright 2017: AgManager.info and K-State Department of Agricultural Economics

