



HAL
open science

Big Data Transformation in Agriculture: From Precision Agriculture Towards Smart Farming

María Angeles Rodríguez, Llanos Cuenca, Angel Ortiz

► To cite this version:

María Angeles Rodríguez, Llanos Cuenca, Angel Ortiz. Big Data Transformation in Agriculture: From Precision Agriculture Towards Smart Farming. 20th Working Conference on Virtual Enterprises (PRO-VE), Sep 2019, Turin, Italy. pp.467-474, 10.1007/978-3-030-28464-0_40 . hal-02478738

HAL Id: hal-02478738

<https://inria.hal.science/hal-02478738>

Submitted on 14 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Big Data Transformation in Agriculture: From Precision Agriculture towards Smart Farming

María Angeles Rodríguez, Llanos Cuenca, and Ángel Ortiz

Research Centre on Production Management and Engineering (CIGIP), Universitat Politècnica de València, Camino de Vera S/N, 46002 València, Spain.
{marodsa4, llcuenca,aortiz}@cigip.upv.es

Abstract. Big data is a concept that has changed the way to analyse data and information in different environments such as industry and recently, in agriculture. It is used to describe a large volume of data (structured or unstructured data), which are difficult to obtain, process or parse using conventional technologies and tools like relational databases or conventional statistics, in a reasonable time for their insight. However, Big Data is applied differently in each area to take advantage of its potential and capabilities. Specially in agriculture that presents more demanding conditions due to its inherent uncertainty, so Big Data methods and models from other environments cannot be used straight away in this area. In this paper, we present a review/update of term Big Data and analyse the evolution and the role of Big Data in agriculture outlined the element of collaboration.

Keywords: Big Data, Smart Farming, Precision Agriculture.

1 Introduction

Agriculture has been a sector with few influences of new technologies, such as Internet of Things (IoT), Cloud Computing or Big Data. Indeed, advanced technologies for data like Big Data or Advanced Data Analysis have not been used until the last decade, although it has always been an area focused on the use and exploitation of data (manual data management in its origins).

Recently, the modern technologies have been introduced in agriculture and it raised new concepts such as Smart Farming. This new concept has burst in to stay and banish old terms like Precision Agriculture. Also, these technologies make the old procedures (re)adapt and new automatized procedures emerge for improve daily management farming and increase collaborative networks between stakeholders.

Therefore, the main contributions of this paper are: the literature update of term Big Data, the proposal of an actual review of Big Data in Agriculture, a description of evolution of Big Data from its origins towards nowadays, and the role of Big Data in collaborative networks.

In order to write this paper, we searched well-known databases, such as Scopus or Elsevier. Searching has focused on most relevant articles to the area of Agriculture

research describing or using Big Data. In this way, we used the following search keywords: ["Precision Agriculture" OR "Smart Agriculture" OR "Precision Farming" OR "Smart Farming"] AND ["Big Data"] Also, the research has been carried out since the year 1980-1990, when the terminology of old term (Precision Agriculture) is already used, until now, when the use of new term (Smart Farming) has spread. Our analysis is divided in two stages: (1) providing a review and an update of term Big Data, and (2) summarizing evolution of Big Data in Agriculture highlighting the aspect of collaboration. Therefore, this paper is organized as follows. In section 2, a review of overview in Big Data are detailed. An evolution of Big Data from Precision Agriculture towards Smart Farming and its implications are described in section 3. Finally, in section 4 conclusions of this paper are exposed.

2 Big data: An Overview

The term Big Data was coined by Cox and Ellsworth in 1997 [1]. And that term was used to describe a problem they observed: "*data sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk*". The supercomputer they used for their Computational Fluid Dynamics research often generated large amounts of data (could exceed 100 Gbytes) that could not visualize or process.

However, the concept Big Data does not refer only to the amount of data. As can be seen in his characterization in 2001, Laney [2] provided Big Data with 3 dimensions:

- *high* Volume (V1): *Size of data*. The important increase in data size.
- *high* Velocity (V2): *Real-Time data*. It has also increased the speed of data generation. This means that their access and analysis must be within a reasonable time to take advantage of this data. That is to say, the data not only must be available immediately, moreover data must be analyzed while data is useful and relevant.
- *high* Variety (V3): *Muti-source*. There is a wide variety of structured or unstructured data sources.

These three dimensions are focused on the storage and processing of data, but there is a lack of analyze the data. Subsequently, in 2012 Laney [3] added two dimensions:

- Value (V4): *Insight*. Extract knowledge from the data.
- Veracity (V5): *High Quality*. The data must be reliable.

The Table 1 quantifies in numerical data the meaning of the 5 V's. In the table it can be noticed that Big Data is more than huge data, also it is different kinds of data, streaming data and data quality. Standing out the significance of last two V's, because they play an important role in decrement corporate earnings due to, for instance, the expenditures for poor data quality and its inadequate management. Specially veracity, because without it; Big Data is less accurate, confidence, consistent or reliable. By this way, inconsistent and unreliable data sets will result in inconsistent and unreliable knowledge [4].

Table 1. Quantification of meaning of the 5 V's [5]

VOLUME	VARIETY	VELOCITY	VALUE AND VERACITY
<i>SIZE OF DATA</i>	<i>MULTI-SOURCE</i>	<i>REAL-TIME DATA</i>	<i>INSIGHT AND HIGH QUALITY</i>
90% of today data has been created in last 2 years	90% of generate data is unstructured	216.000 Instagram posts / every 60 seconds	Poor data quality cost business 600 billion a year
Every day we create 2.5 quintibillion bytes of data	80% of data is video, images and documents	204.000.000 email sent / every 60 seconds	30% of data collected by marketers are not usable
Facebook has 1 petta bytes of storage		50.000 GB/second is the estimated rate of global Internet traffic	

Therefore, as Laney indicates: “*Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization*”.

After Laney, the definition of Big Data was unified and most used the characterization of Big Data in dimensions to define it. Such is the case of NIST [6], IBM [7] and Oracle [8]; which uses 5 V's, only uses 3 V's, and focuses on the challenge of traditional databases, respectively. According to this, the authors' definition is, *Big Data is an effective way to store, process and analyse large amounts of data that are generated nowadays according to business needs*.

The Big Data applications are many, such as healthcare [9] to provide better medical attention to the patient; governments organization [10] to controlled governance dynamics, industry [11] to opens a bright perspective for smart manufacturing or agriculture [4] to improve the productivity in farms or crops.

3 Evolution of Big Data in Agriculture

Agriculture sector was delayed in introducing new technologies such as IoT or Big Data. The IoT concept in agriculture appears for the first time in studies dating from 2010 [12] and it is described as a wireless automation system by monitoring crop and farm data [13]. And Big Data concept in agriculture appears in decade of 2010s and it is described using characterization of 5 V's [14]. Agriculture has always been focused on data. In fact, farmers have always collected amounts of data on their crops and soils for its daily management. That data collection was initially manual and, later, it was automated, until our days is fully automated thanks to the IoT [15].

The use of new technologies in any area, is associated with the appearance of new terms, and in the case of agriculture, two terms must be highlighted: *Precision Agriculture* and *Smart Farming*.

3.1 Precision Agriculture

This concept is special, because over the years it has undergone a change in its meaning and use. Precision Agriculture was presented for the first time in the late 1980s [16], when new Information and Communications Technology (ICT) in agriculture were not yet developed. But it was not until 1990s, when the application of this term began to appear in scientific studies.

At the beginning, in mid 1990s, Precision Agriculture was based as the management of crop, for instance, targeting of fertilizers and herbicides, according to locally determined ('meter-by-meter') requirements, that is to say, precision refers on geolocated data of soil and crop parameters [17]. After ten years, in 2000s, farmer was used complex data such as images or photographs. Detailed spatial information like digitized aerial photographs on crop is needed for controlling crop operations. In this case, Precision Agriculture refers to the application of geospatial technologies (like Geographic Information System, GIS) that allowing farmers access and visualize to images that influence their decisions for the purpose of improving crop performance [18]. Finally, in decade of 2010, as a result to the emergence of IoT and Big Data, everything was based on monitoring data in real time and management of large amount of data that it generates. Furthermore, the concept of Precision Agriculture that is known and used nowadays, is defined as management of agricultural plot by monitoring, data processing and intervention of crops to optimize the consumption of resources and provides the farmers an added value of decision making for exploitation daily operations and management. [19].

The Figure 1 below, shows the evolution of concept Precision Agriculture over the years. As times goes by, it can be observed that in decade of 2000 in agriculture only 2 of the 3 dimensions that were characterized at that time were used. Volume (V1) and Variety (V3) were being used with the introduction of the use of images, because it was possible to have data sources of different types and, in addition, a large size of data was collected. However, Velocity (V2) was not achieved, because data not collected in real time, although sending of the complex data like images were faster than in previous times, it was collected quasi real time not full real time. In fact, it was not until the year 2010, when the Velocity (V2) was achieved, thanks to the monitoring with sensors the sending and acquisition of data in real time.

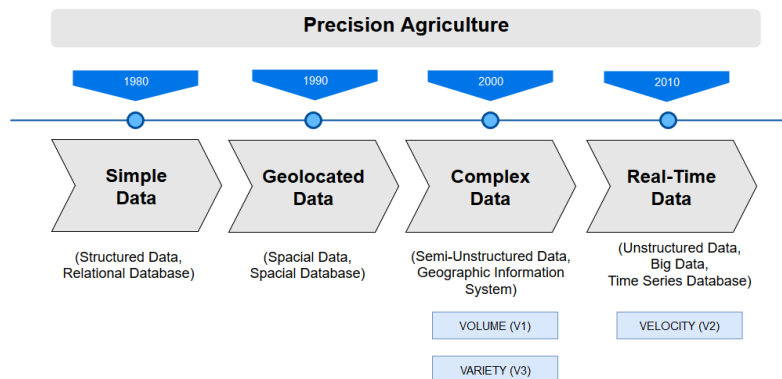


Fig. 1. Evolution of Precision Agriculture

3.2 Smart Farming (Smart Agriculture)

Recently, we have get used to see word *Smart* associate to other words like city (*Smart City*), industry (*Smart Industry*), health (*Smart Health*), etc. The fast proliferation of Smart applications in multiple sectors, is mainly due to the existence of calls for proposals of European Union's Research and Innovation funding programme such as *FP7* (first *Smart* was *Smart City*) or current programme *Horizon 2020* (*Smart Transport, Smart Energy, Smart Farming, Smart Agri-food, etc*).

The term Smart Farming emerges after the concept Precision Agriculture and both are inherently related. At the beginning of 2010s, the term Smart Farming [20] began to be used to refer to that new advanced definition of Precision Agriculture. The aspect of precision (location / geolocation data sent by satellites) disappears, and it materializes on aspect of *Smart* (real-time data sent by smart sensors -IoT architecture-, which highlighted for their ability to communicate its data). In agriculture sector, many Smart Farming applications supports a large amount of heterogeneous real-time data sent by sensors. [21] This kind of data is considered Big Data with 5 V's, because the term Smart assumes the challenges of last two V's: insight and high-quality.

Another aspect relevant in this new term is collaboration. In Precision Agriculture farmer is the main figure or responsible of data and has his own local database system (1980-2000) that he does not share with other farmers, suppliers, etc. New technologies as a Big Data, IoT and Cloud Computing (after 2000), have been able to increase collaboration quickly, because the data is in cloud database systems that can be shared and this real-time data that help to accelerate decisions making. Consequently, decision making be able to include more interlocutors. Everyone can visualize the same data (although each kind of stakeholder sees the information in a different way according to their knowledge). In addition to collaborating in decision making, the information runs between all of them in a connected way, that is, the isolated farm becomes a *connected farm* [22]. The idea of connected farm that appears in Smart Farming, is the fundamental vision of collaborative systems in agriculture. The role played by Big Data in the connected farm is of vital importance and, as already mentioned, it has to assume the challenge of the last two V's. Both daily decisions and strategic decisions of the farm, it is necessary to have adequate information extracted from the data (remember that depending on the stakeholder the information must be presented in a different way) - this is provided by Value -; and also, the information has to be of quality (poor data can cost wrong decisions) -this is provided by Veracity-. And, the role played by IoT in the connected farm is also relevant, mentioned in [15], for instance, FIspace [23] (European business-to-business collaboration IoT platform – belongs to *FP7*) can exchange data or use shared and customizing solutions with minimal costs to help small or medium sized companies.

The Figure 2 below, adds the evolution of concept Smart Farming in Figure 1 previous. It can be observed that in decade of 2010 Smart Farming consumes Precision Agriculture, and it incorporates the last two V's (Value and Veracity). Moreover, this aspect of the transition makes it enhancer faster the collaboration between stakeholders.

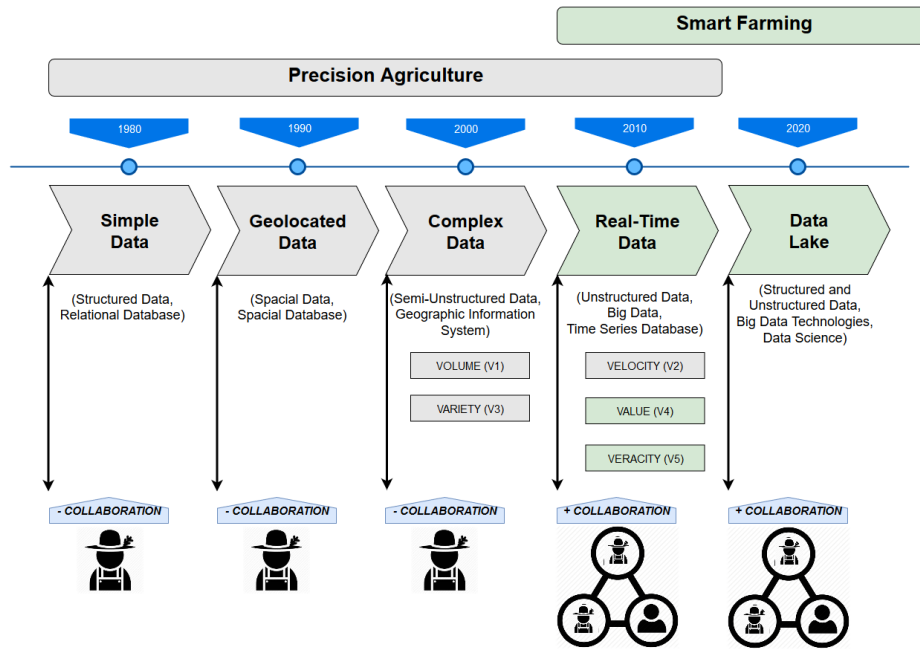


Fig. 2. Evolution of Smart Farming

As an example of Smart Farming application is *Agricolus*. This system uses IoT (crop and soil monitoring), Cloud Computing (process data in cloud), and Big Data (farm management and decision support system for treatments and fertilizers) [24]. And data process of solution is similar to Big Data process. In Big Data, the first step is to capture and save the data -V1, V2 and V3- (in *Agricolus* the recollected data provided by sensors), the second step is to prepare and transform data -V5- (in *Agricolus* the store data is converted in data quality) and, finally, the third step is to analyze the data and act (in *Agricolus* the data is analyzed and visualized in application to support decisions about agricultural operations). Each of the phases of Big Data process presents challenges, mainly the second step (data preparation and transformation). For instance, this step requires 80% of effort and time, due to the particularities of any dataset (different formats, missing values, duplication of data, ...) and its subsequent cleaning, selection and transformation to value data [25]. To reduce this time, could automate data preparation in field of agriculture with a tool such as *START* (Soil daTA Retrieval Tool) [26]. It was designed to automated preparation of soil input data files for multiple crop models. This tool consumed about 33% of time compared with 80% of time in manual preparation. However, they do not mention that automatization must be without lost data quality -V5-.

4 Conclusions

This paper has defined Big Data for characterization of 5 V's (Volume, Velocity, Variety, Value, Veracity) and we have provided a more up-to-date and simple definition of the term. Also, we have updated concepts of Big Data in agriculture, for that purpose we have explained the evolution of new technologies focused on data in agriculture. Finally, we have described collaboration as another aspect of Big Data application.

Without being aware of it, the Big Data approach has always been present in this sector. On a daily basis, large amounts of data are collected and analysed in crops or soils. Before the arrival of digital transformation and TICs, many of these procedures were done manually and only the farmer was responsible for the data and decisions. But nowadays, everything is automated, and the farmer is part of a collaborative network. Moreover, agriculture is to be congratulated, research and innovation in this field is being encouraged from the EU Framework Program called Horizon 2020. For instance, IoF2020 explores the potential of IoT for food and farming industry [27].

To conclude, we have observed that Smart Farming is more collaborative than Precision Agriculture, and only with new technologies as Big Data (5 V's) and IoT, connected farms is a reality. In addition, the future with automatized tools for data preparation or new concept Data Lakes (repository of structured and unstructured big data) present new challenges such as quality and ethics in data and overload of information.

Acknowledgments. All authors acknowledge the partial support of Project 691249, RUC-APS: Enhancing and implementing Knowledge based ICT solutions within high Risk and Uncertain Conditions for Agriculture Production Systems, funded by the EU under its funding scheme H2020-MSCA-RISE-2015; and the project "Development of an integrated maturity model for agility, resilience and gender perspective in supply chains (MoMARGE). Application to the agricultural sector." Ref. GV / 2017 / 025 funded by the Generalitat Valenciana. This first author was supported by the Aid Programme of Research and Development of Universitat Politècnica de València [PAID-01-18].

References

1. Cox, M. and Ellsworth, D. Application-Controlled Demand Paging for out-of-Core Visualization. Proceedings of the 8th Conference on Visualization '97, IEEE Computer Society Press, 1997, p. 235-. (1997)
2. Laney, D. 3D Data Management: Controlling Data Volume, Velocity and Variety. META Group Research Note, 6. (2001)
3. Beyer, M.A. and Laney, D. The Importance of "Big Data": A Definition. Gartner. (2012)
4. Kamilaris, et al. "A Review on the Practice of Big Data Analysis in Agriculture." Computers and Electronics in Agriculture, vol. 143, no. C, Elsevier B.V., 2017, pp. 23–37. (2017)
5. Marr, B. "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read". <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#5671a61d60ba> (2019)

6. NIST. The definition of Big Data. <https://bigdatawg.nist.gov/home.php>
7. IBM. The definition of Big Data. <https://www.ibm.com/analytics/hadoop/big-data-analytics>
8. Oracle. The definition of Big Data. <https://www.oracle.com/big-data/guide/what-is-big-data.html>
9. Shahbaz, M., Gao, Ch., Zhai, L., Shahzad, F., Hu, Y. Investigating the adoption of big data analytics in healthcare: the moderating role of resistance to change. *Journal of Big Data*. 6. 10.1186/s40537-019-0170-y. (2019)
10. Trom L., Cronje J. Analysis of Data Governance Implications on Big Data. In: Arai K., Bhatia R. (eds) *Advances in Information and Communication*. FICC 2019. Lecture Notes in Networks and Systems, vol 69. Springer, Cham (2020)
11. Tao, et al. A Field Programmable Gate Array Implemented Fibre Channel Switch for Big Data Communication towards Smart Manufacturing. *Robotics and Computer Integrated Manufacturing*, vol. 57, Elsevier Ltd, 2019, pp. 166–81. (2019)
12. Lu, Y., Li, X., Zhong, J., Xiong, Y. Research on the innovation of strategic business model in green agricultural products based on Internet of Things (IOT) - May 2010. (2010)
13. Zhao, L., Yin, S., Liu, L., Zhang, Z., Wei, S. A crop monitoring system based on wireless sensor network - December 2011. (2011)
14. Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., and Zhu, Y. Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE*, 104(11), 2207-2219. doi:10.1109/JPROC.2016.2598228. (2016)
15. Rodriguez M.A., Cuenca, L. and Bas, A. FIWARE Open Source Standard Platform in Smart Farming - A Review: 19th IFIP WG 5.5 Working Conference on Virtual Enterprises, PRO-VE 2018, Cardiff, UK, September 17-19, 2018, Proceedings. 10.1007/978-3-319-99127-6_50. (2018)
16. Stafford, J., and LeBars, J. A GPS Backpack System for Mapping Soil and Crop Parameters in Agricultural Fields. *Journal of Navigation*, 49(1), 9-21. (1996)
17. Robert, P.C. Precision agriculture: research needs and status in the USA. In: Stafford, J.V. (Ed.), *Proceedings of the 2nd European Conference on Precision Agriculture*, Part 1. Academic Press, SCI/Sheffield, pp. 19–33. (1999)
18. Long, D. S., Nielsen, G. A., Henry, M. P., and Westcott, M. P. Remote sensing for northern plains precision agriculture. Paper presented at the Space 2000, 208-214. (2000)
19. Ge, Y., Thomasson, J. A., Sui, R. Remote sensing of soil properties in precision agriculture: A review, *Frontiers Earth Sci.*, vol. 5, no. 3, pp. 229–238. (2011)
20. Sundmaeker, H., Verdouw, C., Wolfert, S., and Pérez L. Internet of food and farm 2020. Paper presented at *Digitising the Industry - Internet of Things Connecting Physical, Digital and Virtual Worlds*, River Publishers, Gistrup/Delft, pp. 129-151. (2016)
21. Barmounakis, S., Kaloxylas, A., Groumas, A., Katsikas, L., Sarris, V., Dimtsa, K., Wolfert, S. Management and control applications in Agriculture domain via a FI Business-to-Business platform. *Information Processing in Agriculture*, 2(1), 51–63. (2015).
22. Musat, G., Colezea, M., Pop, F., Negru, C., Mocanu, M., Esposito, C., Castiglione, A. Advanced services for efficient management of smart farms. *Journal of Parallel and Distributed Computing*, Volume 116, 2018, Pages 3-17, ISSN 0743-7315. (2018).
23. FISpace. <https://www.fispace.eu/whatisfispace.html>
24. Agricolus: <https://www.agricolus.com/> (2019)
25. Paton, N. W. Automating data preparation: Can we? should we? must we? *CEUR Workshop Proceedings*, 2324 (2019)
26. Kim, K. S., Yoo, B. H., Shelia, V., Porter, C. H., & Hoogenboom, G. START: A data preparation tool for crop simulation models using web-based soil databases. *Computers and Electronics in Agriculture*, 154, 256-264. doi:10.1016/j.compag.2018.08.023. (2018)
27. IoF2020: <https://www.iof2020.eu/> (2019)