

Chapter 16

Digital Data: Cycle, Standardization, Quality, Sharing, and Security



Antonio Mauro Saraiva, Wilian França Costa, Fernando Xavier, Bruno de Carvalho Albertini, Roberto Augusto Castellanos Pfeifer, Marcos Antonio Simplicio Júnior, and Allan Koch Veiga

16.1 Introduction

The main characteristic of what is called Digital Agriculture, or Agriculture 4.0, is the intensive use of data. It can be said that Digital Agriculture is data-driven. In other words, data, which are becoming increasingly available with spatial and temporal attributes, at high frequencies and on an unprecedented scale, have become essential inputs for the processes that culminate in decision-making.

This digitization phenomenon that is now occurring in agriculture repeats what has already occurred in other areas of human activity, which have been more agile in the incorporation of Information and Communication Technologies (ICTs) in their processes. In fact, the agricultural sector as a whole was slow to adopt these technologies compared to many other sectors and is still one of the least digitized in the world (Krishnan 2017). However, this situation is changing rapidly and, of course, varies greatly among different countries and different sectors of agriculture, including animal production.

It is worth noting that Digital Agriculture has not appeared suddenly, and what we currently see is the result of a long process that began when the first analog electrical monitoring and control systems were incorporated into agricultural tractors and facilities, in the early twentieth century. It gained great momentum with the development of electronics, microelectronics, and ICTs, which occurred from the second half of the twentieth century, as reported by Cox (1997).

A. M. Saraiva (✉) · W. F. Costa · F. Xavier · B. de Carvalho Albertini ·
M. A. Simplicio Júnior · A. K. Veiga
Universidade de São Paulo, Escola Politécnica, São Paulo, Brazil
e-mail: saraiva@usp.br; fxavier@usp.br; balbertini@usp.br; msimplicio@usp.br

R. A. C. Pfeifer
Universidade de São Paulo, Faculdade de Direito, São Paulo, Brazil
e-mail: roberto.pfeiffer@usp.br

From the 1990s, Precision Agriculture (PA) developed remarkably based on the technology already available and was certainly one of the milestones for the transition to Digital Agriculture. This is because PA demands intense use of data acquisition systems in the field, control systems in the machines for the application of inputs at variable rates, and information systems to process an unprecedented amount of data, in spatial and temporal scales. PA boosted digitization, as an important market began to develop and the technology – used widely in other economic sectors – was mature and with more accessible costs for adoption in the field.

After that and again following what was seen in other sectors and industries, there was a very large growth in the development of equipment and systems for data collection, automatic control, information management, and support for decision-making. More recently, a multitude of applications for mobile platforms, smartphones, and tablets to support the most diverse agribusiness activities have also appeared.

Many data-related issues are essentially dependent on the context in which they are inserted and in which they are used. It is thus impossible to detail all those issues for each agricultural use in the space of this text. However, it is important to draw attention to more general aspects that must be considered in any application. In this chapter, some of these aspects are addressed: data life cycle and data science, data standardization, data quality, and data security and legal aspects arising from new data protection laws, recently approved in many countries.

16.2 Big Data, Data Science, and Data Lifecycle

Data is characterized as the “oil” of the digital age.¹ This association is not new: the expression “data is the new oil,” credited to the English mathematician Clive Humby, in 2006, has been widely used to characterize the importance of data in the era of Big Data. This term is used to define the high volume of data, whose collection, storage, circulation, and sharing require specific technology and analytical methods for their transformation into value by companies (EC 2015; Boyd and Crawford 2011).

Traditional approaches to data analysis are not adequate enough in the Big Data era. In addition to the large volume of data, the diversity in the formats and sources and the speed at which they are generated require an evolution in the techniques, methods, and tools. Thus, besides traditional statistical methods, there is a need to use methods from Computer Science to collect, transform, integrate, and analyze data. Data Science meets this demand and is characterized by a multidisciplinary approach in which, in addition to professionals in Statistics and Computer Science, other areas are necessary. In the case of Big Data in agriculture, professionals from the application area are fundamental in projects to extract useful information from agricultural data.

¹The Economist. Available at: <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>

Artificial Intelligence (AI) methods have gained even more importance in the Big Data scenario, both in the search for greater efficiency and for revealing information that would not be obvious only with the application of traditional data analysis methods. However, data analysis goes far beyond the application of these methods and AI, being a part of the data life cycle that contains different activities, from planning to analyzing and producing results, as illustrated in Fig. 16.1.

This data cycle illustrates, generally speaking, data management activities, both in a scientific context and a business context. Although each stage of this cycle can be performed by a different professional, the importance of the multidisciplinary vision brought by Data Science is highlighted.

Experts in the application domain, in this case, experts in agriculture (in a broad sense), have a fundamental role from the very planning stage, in which the objectives in the use of these data will be defined. Therefore, before defining what data to collect, it is necessary to establish the objectives and sources. However, the participation of domain experts is not limited to the planning stage. In fact, it can be and often is also related to other activities in the cycle defined in Fig. 16.1. This reinforces the importance that professionals linked to the agricultural sector – agronomists, zoo-technicians, agricultural engineers, agricultural technicians, etc. – increasingly have training that allows them to interact with digital technology professionals.

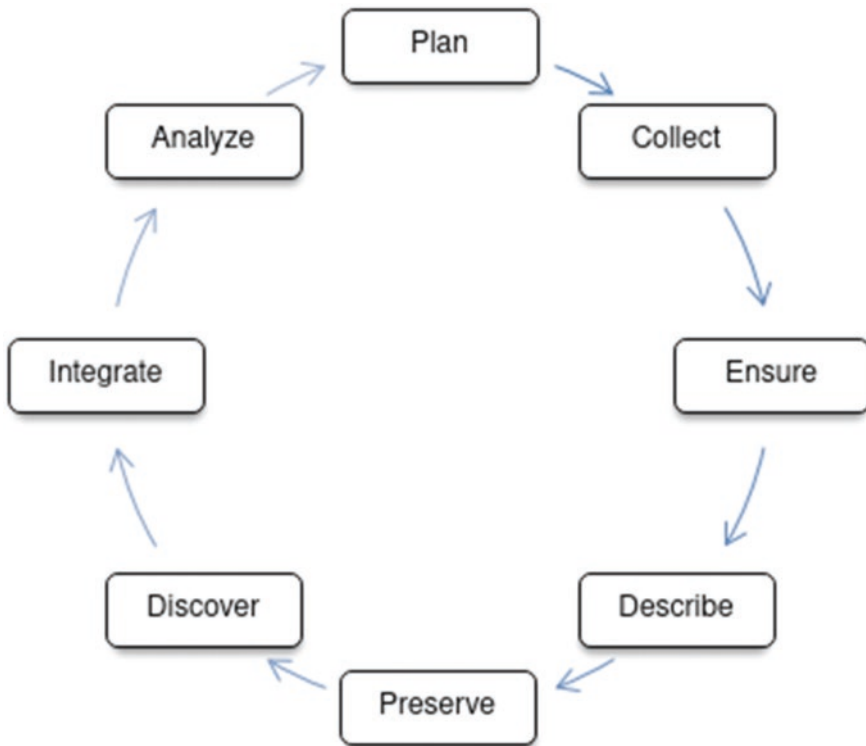


Fig. 16.1 Data life cycle model. (Source: Adapted from Silva 2017)

16.3 Standardization of Data and Communication in Agriculture

Agricultural data are quite diverse from the point of view of their sources and formats and do not usually follow a widely accepted standardization. This stems from the very characteristic of the sector, which is geographically very dispersed and technologically very uneven. Data standardization is one of the key factors for the success of digital agriculture. It allows two entities (software, people, institutions, etc.) to exchange data that will be interpreted and treated in the same way, regardless of physical or temporal distance, avoiding errors and reducing the costs related to data conversion.

Data standardization is a set of collaborative documents that indicate the consensus of a specific community on the representation, format, definition of meaning, structuring, marking, transmission, manipulation, use, and management of data. Below are the main benefits of standardization:

- It is described by qualified people. When an entity sees the need for standardization or when the volume of data on a particular subject increases significantly, interested people are brought together in public calls, in which the definitions behind a standardization are discussed. These groups are made up of multiple profiles, from academic, business, government, and local producers.
- It allows for better transparency and a homogeneous understanding of the data. Standardizations often have a related vocabulary, which means that the understanding of a concept is also standardized, that is, two different entities will understand the data in a similar way.
- Saves resources. Although the adoption of standards is often more expensive in relation to non-standard data, in the long run, that cost is often paid since tools, codes, methods, and resources can be reused without the need to adapt them. Standardized data are said to have a longer lifetime than non-standardized data.

Most standards are written and formalized by entities with a reputation or mandate in certain areas, such as associations, governments, or professional societies. Through these entities, some data and communication standards for agriculture have been developed, being at different stages of maturity and adoption. If you have questions about which standard you should use, please contact your local standardization body, for example, ARSO (Africa), ABNT/Brazil, SCC/CSA (Canada), SAC (China), AFNOR (France), DIN (Germany), UNI (Italy), JISC (Japan), BSI (UK), or ANSI (USA). ISO (International Organization for Standardization, <https://www.iso.org/>) can be contacted to point out your local standardization body.

16.3.1 *AgroXML*

AgroXML (<http://www.agroxml.de/>) is one of the main standards in the field of agriculture since it covers a wide range of topics, from precision agriculture to the food production chain or the management of smart agricultural companies. It is

based on the XML standard (eXtensible Markup Language), whose main features are ease of use and extensive support. The standard was developed independently by a study group at the University of Hohenheim (Stuttgart, Germany) in 2004. However, the flexibility resulting from the choice of XML has spread the standard worldwide, which is one of the main data standards used. Note that AgroXML is open-source, and its use is free.

16.3.2 AgMES

The AgMES (<http://aims.fao.org/standards/agmes/>, Agricultural Metadata Element Set) was developed to organize information from the agricultural area, including any digital information. Today, it is maintained by the Agricultural Information Management Standards (<http://aims.fao.org/>) of the United Nations Food and Agriculture Organization (FAO), but its development is stagnant. New adoptions of the standard are discouraged, but it is still useful for organizers of information (e.g., libraries, collections, etc.). For new adoptions, it is suggested to use AGRIS.

16.3.3 AGRIS

AGRIS (<http://aims.fao.org/agris-network>), also maintained by FAO, is the successor to AgMES. It aims to catalog information on food production in general, with a focus on agriculture (including precision agriculture). The standard is predominantly bibliographic, with significant use in academia, but little commercial adoption, which is why we will not extend its description.

16.3.4 AGROVOC

AGROVOC (<http://aims.fao.org/standards/agrovoc/>) is another mechanism maintained by FAO, consisting of extensive multilingual vocabulary (including English and Portuguese) on all aspects of food production. It is used as a guide for data storage and communication. Using AGROVOC, data interpretation should be clear and unambiguous, as those involved have a unique source of concepts, terms, and relationships between them. The standard covers about 36,000 concepts, which makes it the main semantic reference for agricultural standards. Its use is free through a Creative Commons license.

16.3.5 *ISOBUS*

ISOBUS is the main standard in the agricultural area for use in machinery. It is a communication standard used for data interoperability of machines and implements (M2M). Its adoption allows a producer to purchase a machine from one manufacturer and service or implement from another, for example, as long as both are adherent to the standard.

This standard consists of a high-resilience serial communication based on CANBus, the communication network used in non-agricultural vehicles, such as automobiles and trucks. It involves the communication protocol (physical and logical data format), the user interface in the machines (e.g., the on-board computer on the tractor), control of operations, and file servers. It also has a linked vocabulary, which allows data collected through ISOBUS to be interpreted in the same way by tools from different manufacturers. The main advantage of ISOBUS is the economy, since, by adopting it, the producer avoids redundancy in the equipment and can reuse or connect machines and implements from different suppliers. However, as message exchanges and file formats are also standardized, the standard has also been used in data analysis. It is an international standard (ISO 11783) broadly adopted by the agricultural machinery industry worldwide. It is strongly recommended that any implement or machine purchased be adherent to the standard.

16.3.6 *Open Geospatial Consortium Standards (OGC®)*

We cannot discuss data standardization in agriculture without referring to the standards specified by the Open Geospatial Consortium (OGC). The OGC is an international consortium of more than 530 companies, government agencies, research organizations, and universities geared toward making geospatial information and services (localization) Findable, Accessible, Interoperable, and Reusable (the so-called FAIR principles) (Wilkinson et al. 2016). Created in 1994, the consortium was aimed at standardizing data used in Geographic Information Systems (GIS). Today, OGC operates in the development and implementation of open standards for geospatial content and services, sensor networks, the Internet of Things, georeferenced data processing, and data sharing (Mckee 2020).

Several OGC standards have become popular over the years, as the use of geospatial data evolved. Web service standards, such as the Web Map Service ((WMS) vector map bitmap rendering service), Web Feature Service (service for accessing and editing geometries (vector data)), and the Web Coverage Service (service for data rendering raster), are currently supported on all GIS tools and servers in the market. Map servers, such as Geoserver (<http://geoserver.org/>), MapServer (<https://www.mapserver.org/>), and the adoption of these standards in GIS tools already consolidated in the market, such as ESRI ArcGIS Server (<https://www.esri.com/en-us/>

[arcgis/products/arcgis-enterprise/overview](#)), demonstrate the success in using these standards, in different communities and with different focuses.

Currently, OGC has working groups directly concerned with the development of specific standards related to agricultural resources. We here highlight the **Agriculture Domain Working Group** (DWG), some purposes of which are (Di and Charvat 2020): to examine and propose the possibilities of aligning and harmonizing agricultural information exchange standards between initiatives and organizations, such as CEFACT (UN), ISO TC 23, ISOBus, AgroXML, OGC, W3C, etc.; and the development of a reference architecture for the use of coding standards and the OGC interface in common agricultural activities.

16.4 AgGateway Standards

AgGateway is a global nonprofit organization, whose members develop standards and other resources so that agricultural companies can access information quickly by adopting standards for interoperability, facilitating the transition to digital and sustainable agriculture (<https://www.aggateway.org/AboutUs/Mission.aspx>). The Ag Data Application Programming Toolkit (ADAPT) consists of an Agricultural Application Data Model, a common API (Application Programming Interface), and a combination of proprietary and open-source data conversion plug-ins. Companies that market Agricultural Management Information Systems are responsible for building their own implementation of mapping the Agricultural Application Data Model to their specific data model. They include several standards (Ferreira 2017), such as for irrigation data (PAIL – Precision Ag Irrigation Language) (Aggateway 2020a), the integration of planting data and the use of fertilizers (SPADE – Fertilization Data Standards) (Aggateway 2020b), and semantic image identification for agricultural remote sensing in GeoTIFF format (PICS – Imagery Tagging) (Ferreira 2019).

16.5 Data Quality

Data quality has significant consequences and effects on its use. Poor data quality is estimated to cause 8–12% of revenue loss and to represent 40–60% of companies' service costs (Redman 1998).

In the agricultural sector, there are great benefits in assuring the quality of the data used by specialists for decision-making and for supporting data-dependent activities, such as income forecasting, monitoring, and planning (Malaverri and Medeiros 2012). Therefore, the assessment and management of data quality for the improvement of “data fitness-for-use” are actions justified by cost reduction and by making more assertive and efficient decisions.

16.6 Quality Assessment

In general, management methodologies presuppose clear and objective definitions of success metrics. In the field of data quality, it is not different. Improving this quality depends on clear metrics of success, the so-called data quality dimensions. Therefore, there is a consensus in the literature that data quality is a multidimensional concept, and its meaning in a project is defined by a series of relevant dimensions in a given context, such as consistency, precision, accuracy, completeness, trust, reputation, accessibility, among others (Wang and Strong 1996).

Identifying which data quality dimensions are relevant to the success of a project, measuring the quality in these dimensions, and establishing criteria to assess whether that quality is fit-for-use in a given context are essential for efficient data quality management (Veiga et al. 2017).

16.7 Quality Management

The objective of this management is to improve the quality of the data by the prevention and correction of errors that directly or indirectly degrade the quality of the data in one or more relevant dimensions of a project.

Two strategies can be adopted for this: quality control and quality assurance. The first seeks to optimize the measures of the data quality dimensions whenever possible, without losing data but assuming that the quality may not be fully in accordance with the design criteria. Conversely, quality assurance assumes that the data have a quality level totally in accordance with the project criteria, which may imply data loss if a subset of data does not meet the pre-established criteria (Veiga et al. 2017).

Regardless of the strategy adopted over time in the organization, the data used to support punctual decision-making must be evaluated with quality measures (e.g., precision, accuracy, consistency, completeness), according to the pre-established criteria in the context of the project.

16.8 Data Sharing and Security

Many companies have been increasing their profits by using information about users' behavior, preferences, needs, expectations, desires, and opinions. Several of the innovative businesses from the digital age rely on user data that are quite often shared among different platforms. For this reason, many already consider data as a commodity (Morando et al. 2014). Data collection, analysis, and customization are also part of agribusiness, whereby it is used to improve products, increase sales, or learn about consumer preferences and adapt to them (through advertising). For this

reason, some authors highlight the existence of a data value chain (Curry 2016; Miller and Mork 2013) in agribusiness.

In this scenario where data sharing takes a major role, it is important to analyze which security aspects are relevant and which actions should be taken to ensure compliance with data protection laws, such as the European General Data Protection Regulation (GDPR) or the Brazilian General Data Protection Law (*Lei Geral de Proteção de Dados – LGPD*).

16.9 Data Security

When allowing data to be shared among different institutions, it is worth considering some essential aspects of information security. Typically, a robust system should provide a combination of the following basic security pillars:

- *Confidentiality or secrecy* – Only authorized users can access information transmitted or stored in the system. For example, personal data of customers and suppliers should never be transferred in cleartext over the Internet, but instead use secure communication technologies (e.g., in the case of a web page, HTTPS should be used instead of HTTP). This approach aims not only at protecting the privacy of any entity whose data is shared but also to avoid any possible competitive damage resulting from strategic information leakages.
- *Integrity* – If some piece of information is modified without authorization, whether accidentally or on purpose, it must be possible to detect this modification. Note that it is not always possible to prevent or undo the change, but it is essential to enable detection aiming to prevent misguided actions, taken after the analysis of bogus data. For example, when using sensors to continuously monitor soil quality, the integrity of the collected data must be protected against modification; otherwise, this might lead to the application of an undue amount of fertilizers, possibly compromising the entire crop.
- *Authenticity* – During the whole duration of a communication, both sender and receiver should be able to identify each other's messages. This service is particularly important to prevent intrusion attempts, such as an attacker trying to impersonate a legitimate user of the system, in an attempt to access sensitive data; the insertion of malicious sensor nodes in a field monitored by an IoT system, aiming to inject false information into the system and, as a result, manipulate its actions.
- *Non-repudiation* – Guarantee that a user cannot deny having created or sent a message. This security service is directly related to the concept of digital signatures: when signing a document, one cannot subsequently deny such a signature, so the document author can be held responsible in case of misconduct. For this reason, the deployment of non-repudiation services into a system is usually a requirement for constructing robust audit mechanisms.

- *Availability* – Legitimate users of the system should not be prevented from accessing it. A common example of an attack against the availability of systems is the so-called Denial of Service (DoS) attack. Several mechanisms can be used to mitigate or lessen the impact of such threats. An example is to filter messages sent by automatic means, for example, by requiring the solution of a challenge that is not easily solved by robots (a mechanism commonly known as “CAPTCHA”). When it is necessary to support such automation, it is common to require senders to be authenticated, and then limit their transmission rate to avoid attempts to monopolize the system resources. Another common approach for dealing with the threat of overload involves the adoption of some degree of system redundancy and elasticity, which is commonly achieved by employing cloud computing technologies.

To identify which security services are a priority in each application scenario, it is important to consider the characteristics of the system and of the data it handles. For example, consider an automated irrigation system in which soil moisture is constantly monitored and sprinklers are activated whenever necessary. In this case, data integrity, authenticity, and availability are likely more important than confidentiality and non-repudiation; after all, avoiding excess or lack of water is more relevant than preventing third parties from accessing moisture-related readings, or proving that a given sensor was responsible for sending a specific reading. However, when the data transferred involves product prices, the result of negotiations, the contents of contracts, or other strategic information, all of the hereby listed security services can become similarly relevant.

Finally, it is worth noting that there are technologies created specifically to facilitate the task of securely sharing data among different organizations, taking into account the aforementioned security concerns. One particularly prominent solution is OAuth, or Open Authorization (<https://oauth.net/2/>), an open security protocol that enables data owners to delegate access to (part of) their data to a third party, for a specific time. The protocol is currently in version 2.0, and it is used in the construction of several solutions in which the control of the information flow is centered on users rather than on the servers employed for storing users’ data.

16.10 Protection of Personal Data

Most countries exporting and importing agricultural products have personal data protection legislation. This is the example of Brazil and its Data Protection General Law, and the countries of the European Union whose General Data Protection Regulation, GDPR, was enacted in May 2016.

Thus, the collection and sharing of personal data of individuals involved in agricultural production must comply with the rules imposed by the applicable data protection legislation.

The best-known hypothesis of legal authorization is the consent of the data subject (Article 6, 1, of the GDPR). In this context, for example, if the agricultural producer wishes to collect information about the shopping habits of his consumers and share such data with third parties, he will have to obtain prior consent from the acquirers or, alternatively, anonymize such data.

Notwithstanding, there are other possibilities for using personal data that do not depend exclusively on the consent of the data subject.

Some examples can be given concerning the agribusiness sector. There are cases in which data collection and transmission to inspection bodies are mandatory (e.g., pesticide user data and the place where the product will be applied). In these cases, data processing is allowed regardless of the prior consent of the data subject (Article 6°, 1, c of the GDPR).

Another hypothesis occurs when the contract established between the consumer and the agricultural company requires certain data to be collected and shared. This would be the case, for example, of the farmer who buys seeds from a certain producer. There may be a need to collect and process the consumer's personal data to enable the contract conclusion. In this case, the collection and sharing of data between seller and producer do not require the prior consent of the data subject, because it is permitted by the contract.

However, clear and adequate information on the processing of the data collected will always be necessary, whatever the basis of the lawfulness of processing.

16.11 Final Considerations

Data is a new source of wealth, a precious asset for those who generate it and for everyone involved in the chain of its use, symbolized by the data life cycle. It is a fact already well known and explored in various industries and, currently, also in agriculture. Data is the basic input for obtaining information (i.e., contextualized data) and knowledge, which supports decision-making and the formulation of business and government policies.

However, there are countless aspects to be considered for the entire cycle to develop effectively in the best possible way, especially considering an area as complex as agriculture. This is an area where data is generated in a very distributed way, in time and space, by a huge variety of users, devices, equipment, and systems that are very heterogeneous from various points of view, including technological and cultural, among others.

In this chapter, some of these aspects were presented that apply in general since a specific approach for each use would not make sense in the scope of this text. With that, we tried to offer a first approach, which should be further explored by the reader. It is worth mentioning that, with the growing importance of data from current businesses, Data Science has emerged as an important area, which corroborates the fact that, in the space of a chapter, the possible approach is the introduction of the subject.

There is no doubt, however, that data is also one of the main assets of the agricultural sector and the main foundation of Digital Agriculture.

Abbreviations/Definitions

- AI: Artificial intelligence is the intelligence demonstrated by machines, as opposed to the natural intelligence displayed by animals including humans.

Take Home Message/Key Points

- The main characteristic of what is called Digital Agriculture or Agriculture 4.0 is the intensive use of data.
- Digital Agriculture is data-driven. Data, which are becoming increasingly available with spatial and temporal attributes, at high frequencies and on an unprecedented scale, have become essential inputs.
- Data must be considered in regard to its life cycle, standardization, quality, security, and legal aspects to be used in its full benefits to farming.

References

- Aggateway (2020a) Precision Ag Irrigation Language (PAIL). Available at [https://www.aggateway.org/eConnectivityActivities/Implementation/PrecisionAgIrrigationLanguage\(PAIL\).aspx](https://www.aggateway.org/eConnectivityActivities/Implementation/PrecisionAgIrrigationLanguage(PAIL).aspx). Accessed 12 Mar 2020
- Aggateway (2020b) Specific Initiatives. Available at <https://www.aggateway.org/GetConnected/SpecificInitiatives.aspx>. Accessed 13 Mar 2020
- Boyd D, Crawford K (2011) Six provocations for Big Data. In: Symposium on the dynamics of the internet and society: a decade in internet time, September. Available at <https://ssrn.com/abstract=1926431> or <https://doi.org/10.2139/ssrn.1926431>. Accessed 13 Mar 2020
- Cox SWR (1997) Measurement and control in agriculture. Blackwell Science, London. ISBN 0-632-04114-5
- Curry E (2016) The big data value chain: definitions, concepts, and theoretical approaches. In: Cavanillas J, Curry E, Wahlster W (eds) New horizons for a data-driven economy. Springer, Chain
- Di LP, Charvatdim F, Sakuda LO (org) (2020) Radar AgTech Brasil 2019: mapeamento das startups do setor agro brasileiro. Brasília: Embrapa, SP Ventures e Homo Ludens, K. Agriculture DWG. Available at <https://www.ogc.org/projects/groups/agriculturedwg>. Accessed 13 Mar 2020
- EC – European Commission (2015) Towards a thriving data-driven economy. Off J Eur Union 58:61–65
- Ferreira R (2017) ADAPT Public Space. 15 set. Available at <https://www.aggateway.org/GetConnected/AgGateway%E2%80%99sADAPT.aspx>. Accessed 14 Mar 2020
- Ferreira, R (2019) A AgGateway Post-Image Collection Specification (PICS). Available at <https://aggateway.atlassian.net/wiki/x/XABrDw>. Accessed 13 Mar 2020
- Krishnan N (2017) Cultivating Ag Tech: 5 trends shaping the future of agriculture. Available at <https://www.cbinsights.com/research/agtech-startup-investor-funding-trends>. Accessed 14 Mar 2020
- Malaverri JEG, Medeiros CB (2012) Data quality in agriculture applications. In: GEOINFO, 13 November 25–27 de 2012, Campos do Jordão, SP, Brazil. Proceedings of the Brazilian Symposium on GeoInformatics. Campos do Jordão, SP, pp 128–139
- Mckee L (2020) OGC history (detailed). Available at <https://www.ogc.org/ogc/historylong>. Accessed 13 Mar 2020

- Miller HG, Mork P (2013) From data to decisions: a value chain for big data. In *It Profess* 15:57–59. <https://doi.org/10.1109/MITP.2013.11>
- Morando F, Iemma R, Raitieri E (2014) Privacy evaluation what empirical research on users' valuation of personal data tells us. *Internet Pol Rev* 3(2). <https://doi.org/10.14763/2014.2.283>
- Redman T (1998) The impact of poor data quality on the typical enterprise. *Communications of the ACM*. <https://doi.org/10.1145/269012.269025>
- Silva DL (2017) Estratégia computacional para apoiar a reprodutibilidade e reuso de dados científicos baseado em metadados de proveniência. PhD thesis – Escola Politécnica, Universidade de São Paulo, São Paulo. <https://doi.org/10.11606/T.3.2017.tde-05092017-095907>. Available at <https://www.teses.usp.br>. Accessed 3 Jan 2020. (English title: Computational strategy to support reproducibility and reuse of scientific data based on provenance metadata)
- Veiga AK, Saraiva AM, Chapman AD, Morris PJ, Gendreau C, Schiegel D, Robertson TJ (2017) A conceptual framework for quality assessment and management of biodiversity data. *PLOS One* 12:e0178731. Available at <https://doi.org/10.1371/journal.pone.0178731>. Accessed 3 Jan 2020
- Wang RY, Strong DM (1996) Beyond accuracy: what data quality means to data consumers. *J Manag Inf Syst* (Published by M.E. Sharpe, Inc.) 12:5–33. Available at <http://www.jstor.org/stable/40398176>. Accessed 3 Jan 2020
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, Hoen Peter AC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, Lei J, Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 3:160018. Available at <https://doi.org/10.1038/sdata.2016.18>. Accessed 3 Jan 2020