Chapter 1:

Big Data for the Greater Good: An Introduction

Vincent Charles[1] **and Ali Emrouznejad**[2]

[1] Buckingham Business School, University of Buckingham, Buckingham, UK

v.charles@buckingham.ac.uk

[2] Aston Business School, Aston University, Birmingham, UK

a.emrouznejad@aston.ac.uk

Abstract Big Data, perceived as one of the breakthrough technological developments of our times, has the potential to revolutionize essentially any area of knowledge and impact on any aspect of our life. Using advanced analytics techniques such as text analytics, machine learning, predictive analytics, data mining, statistics, and natural language processing, analysts, researchers, and business users can analyze previously inaccessible or unusable data to gain new insights resulting in better and faster decisions, and producing both economic and social value; it can have an impact on employment growth, productivity, the development of new products and services, traffic management, spread of viral outbreaks, and so on. But great opportunities also bring great challenges, such as the loss of individual privacy. In this chapter, we aim to provide an introduction into what Big Data is and an overview of the social value that can be extracted from it; to this aim, we explore some of the key literature on the subject. We also call attention to the potential 'dark' side of Big Data, but argue that more studies are needed to fully understand the downside of it. We conclude this chapter with some final reflections.

Cite this chapter as:

Charles V., Emrouznejad A. (2019) Big Data for the Greater Good: An Introduction.

In: Emrouznejad A., Charles V. (eds) Big Data for the Greater Good. Studies in Big Data,

vol 42. Springer, Cham

https://doi.org/10.1007/978-3-319-93061-9_1

ISBN: 978-3-319-93060-2 (p) and 978-3-319-93061-9 (o)

1 The Hype Around Big Data and Big Data Analytics

- Is it possible to predict whether a person will get some disease 24 hours before any symptoms are visible?
- Is it possible to predict future virus hotspots?
- Is it possible to predict when or where a fraud or crime will take place before it actually happens?
- It is possible to predict traffic congestion up to three hours in advance?
- Is it possible to predict terrorists' future moves?
- Can we support in a better way the wellbeing of people?

These are some of the questions that the Big Data age promises to have an answer for. We should note, at the outset, that some of these questions have already been answered, fostering new waves of creativity, innovation, and social change. But the true potential of Big Data in any of the areas mentioned is yet to be fully unlocked.

Big Data is a big phenomenon that for the past years has been fundamentally changing not only what we know, but also what we do, how we communicate and work, and how we cooperate and compete. It has an impact at the individual, organizational, and societal level, being perceived as a breakthrough technological development (Fichman, Dos Santos, & Zheng, 2014).

Today, we are witnessing an exponential increase in 'raw data', both human and machine-generated; human, borne from the continuous social interactions and doings among individuals, which led McAfee and Brynjolfsson (2012) to refer to people as "walking data generators" (p. 5); machine-generated, borne from the continuous interaction among objects (generally coined the 'Internet-of-Things'), data which is generally collected via sensors and IP addresses. According to Baesens *et al*. (2016), Big Data comes from five major sources (see Figure 1 for a visual representation):

- *Large-scale enterprise systems*, such as enterprise resource planning, customer relationship management, supply chain management, and so on.
- *Online social graphs*, resulting from the interactions on social networks, such as Facebook, Twitter, Instagram, WeChat, and so on.
- *Mobile devices*, comprising handsets, mobile networks, and internet connection.
- *Internet-of-Things*, involving the connection among physical objects via sensors.

- *Open data/ public data*, such as weather data, traffic data, environment and housing data, financial data, geodata, and so on.

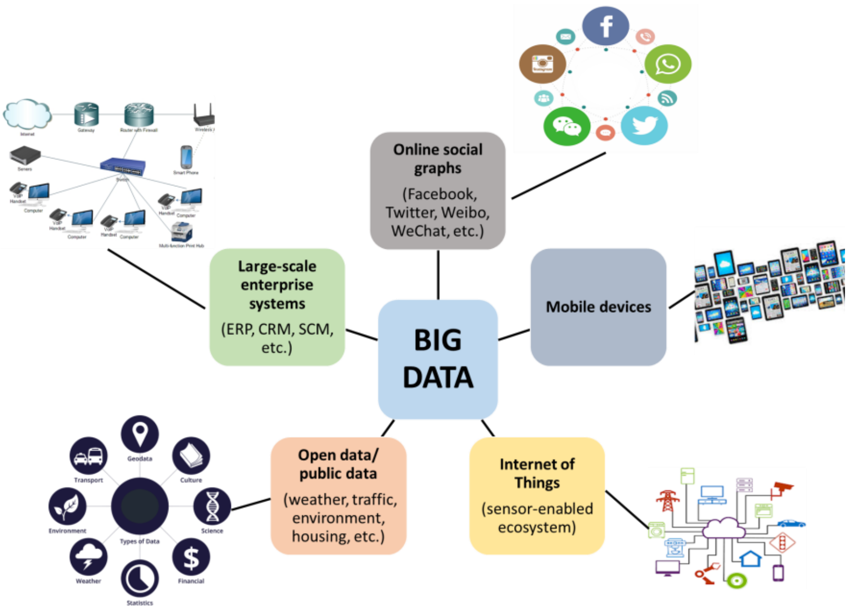

Figure 1. Sources of Big Data. Adapted from Baesens *et al.* (2016).

Advanced analytics techniques such as text analytics, machine learning, predictive analytics, data mining, statistics, and natural language processing, are just as important as the Big Data itself. This may sound somehow obvious and trivial, but it is important to be clear about the weight that each holds in the discussion. On the one hand, were it not for the possibility to collect the large amounts of data being generated, the development of advanced analytics techniques would be irrelevant. On the other hand, the availability of the huge data would mean nothing without advanced analytics techniques to analyze it. Advanced analytics techniques and Big Data are, thus, intertwined. And thanks to Big Data Analytics, we now have the possibility to transform all of the data into meaningful information that can be explored at various levels of the organization or society (Charles, Tavana, & Gherman, 2015).

It is useful to distinguish among three different types of analytics: descriptive, predictive, and prescriptive. It should be noted that today, nonetheless, most of the efforts are directed towards predictive analytics.

- *Descriptive analytics* help answer the question: *What happened?* It uses data aggregation and data mining (dashboards/scorecards and data visualization, among others). It helps to summarize and describe the past and is useful as it can shed light onto behaviors that can be further analyzed to understand how they might influence future results. For example, descriptive statistics can be used to show average pounds spent per household or total number of vehicles in inventory.

- *Predictive analytics* help answer the question; *What will or could happen?* It uses statistical models and forecasts techniques (regression analysis, machine learning, neural networks, golden path analysis, and so on). It helps to understand the future by means of providing estimates about the likelihood of a future outcome. For example, predictive analytics can be used to forecast customer purchasing patterns or customer behavior.

- *Prescriptive analytics* help answer the question: *How can we make it happen? Or what should we do?* It uses optimization and simulation algorithms (algorithms, machine learning, and computational modelling procedure, among others). It helps to advise on possible outcomes before decisions are made by means of quantifying the effect of future decisions. For example, prescriptive analytics can be used to optimize production or customer experience.

Despite the increased interest in exploring the benefits of Big Data emerging from performing descriptive, predictive, and/or prescriptive analytics, however, researchers and practitioners alike have not yet agreed on a unique definition of the concept of *Big Data* (Beyer & Laney, 2012; Charles & Gherman, 2012; Hammond, 2013; Laney, 2001; and Ohlhorst, 2013). We have performed a Google search of the most common terms associated with Big Data and Figure 2 depicts these terms that we have compiled from across the big data spectrum.

Figure 2. A broad view of the Big Data terminology.

The fact that there is no unique definition of Big Data is not necessarily bad, since this allows the possibility to explore facets of Big Data that may otherwise be constrained by a definition; but, at the same time, it is surprising if we consider that what was once considered to be a problem, that is, the collection, storage, and processing of the large amount of data, today is not an issue anymore; as a matter of fact, advanced analytics technologies are constantly being developed, updated, and used. In other words, we have the IT technology to support the finding of innovative insights from Big Data (Davenport, Barth, & Bean, 2012), while we lack a unified definition of Big Data. The truth is that indeed, this is not a problem. We may not all agree on *what* Big Data is, but we do all agree that Big Data *exists*. And *grows* exponentially. And this is sufficient because what this means is that we now have more information than ever before and we have the technology to perform analyses that could not be done before when we had smaller amounts of data.

Let us consider, for example, the case of WalMart, who has been using predictive technology since 2004. Generally, big retailers (WalMart included) would collect information about their transactions for the purpose of knowing how much they are selling. But WalMart took advantage of the benefits posed by Big Data, and took that extra step when it started analyzing the trillions of bytes' worth of sales data, looking for patterns and correlations. One of the things they were able to determine was what customers purchase the most ahead of a storm. And the answer was Pop-Tarts ("Why WalMart Always Stocks Up On Strawberry Pop-Tarts Before a Hurricane", 2017), a sugary pastry that requires no heating, lasts for an incredibly long period of time, and can be eaten at any meal. This insight allowed

WalMart to optimize its supply of Pop-Tarts in the months or weeks leading up to a possible storm. The benefit for both the company and the customers is, thus, obvious.

Considering the above example, one of the conclusions we can immediately draw is that the main challenge in the Big Data age remains how to use the newly-generated data to produce the greatest value for organizations, and ultimately, for the society. Einav and Levin (2013, p.2) captured this view when they elegantly stated that Big Data's potential comes from the "identification of novel patterns in behaviour or activity, and the development of predictive models, that would have been hard or impossible with smaller samples, fewer variables, or more aggregation". We thus also agree with the position taken by Baesens *et al.* (2016, p. 808), who stated that "analytics goes beyond business intelligence, in that it is not simply more advanced reporting or visualization of existing data to gain better insights. Instead, analytics encompasses the notion of going beyond the surface of the data to link a set of explanatory variables to a business response or outcome".

It is becoming increasingly clear that Big Data is creating the potential for significant innovation in many sectors of the economy, such as science, education, healthcare, public safety and security, retailing and manufacturing, e-commerce, and government services, just to mention a few – we will discuss some of this potential later in the chapter. For example, according to a 2013 Report published by McKinsey Global Institute, Big Data analytics is expected to generate up to $190 billion annually in healthcare cost savings alone by 2020. Concomitantly, it is also true that despite the growth of the field, Big Data Analytics is still in its incipient stage and comprehensive predictive models that tie together knowledge, human judgement and interpretation, commitment, common sense and ethical values, are yet to be developed. And this is one of the main challenges and opportunities of our times. The true potential of Big Data is yet to be discovered.

We conclude this section with the words of Watson (2014, p. 1247), who stated that "The keys to success with big data analytics include a clear business need, strong committed sponsorship, alignment between the business and IT strategies, a fact-based decision-making culture, a strong data infrastructure, the right analytical tools, and people skilled in the use of analytics."

2 What is Big Data?

The term 'Big Data' was initially used in 1997 by Michael Cox and David Ellsworth, to explain both the data visualization and the challenges it posed for computer systems (Cox & Ellsworth, 1997). To say that Big Data is a new thing is to some extent erroneous. Data have always been with us; it is true, however, that during the 1990s and the beginning of the 2000s, we experienced an increase in

IT-related infrastructure, which allowed to store the data that was being produced. But most of the time, these data were simply that: stored – and most probably forgotten. Little value was actually being extracted from the data. Today, besides the required IT technology, our ability to generate data has increased dramatically – as mentioned previously, we have more information than ever before, but what has really changed is that we can now analyze and interpret the data in ways that could not be done before when we had smaller amounts of data. And this means that Big Data has the potential to revolutionize essentially any area of knowledge and any aspect of our life.

Big Data has received many definitions and interpretations over time and a unique definition has not been yet reached, as indicated previously. Hammond (2013) associated Big Data with evidence-based decision-making; Beyer and Laney (2012) defined it as high volume, high velocity, and/or high variety information assets; and Ohlhorst (2013) described Big Data as vast data sets which are difficult to analyze or visualize with conventional information technologies. Today, it is customary to define Big Data in terms of data characteristics or dimensions, often with names starting with the letter ‘V’. The following four dimensions are among the most often encountered (Laney, 2001):

Volume: It refers to the large amount of data created every day globally, which includes both simple and complex analytics and which poses the challenge of not just storing it, but also analyzing it (Laney, 2001). It has been reported that 90% of the existent data has been generated in the past two years alone (IBM, 2016). It is also advanced that by 2020, the volume of data will be 40 zettabytes, 300 times bigger than the volume of data in 2005 (Herschel & Mori, 2017).

Velocity: It refers to the speed at which new data is generated as compared to the time window needed to translate it into intelligent decisions (Laney, 2001). It is without doubt that in some cases, the speed of data creation is more important than the volume of the data; IBM (2013) considered this aspect when they stated that “for time-sensitive processes such as catching fraud, big data must be used as it streams into your enterprise in order to maximize its value”. Real-time processing is also essential for businesses looking to obtain a competitive advantage over their competitors, for example, the possibility to estimate the retailers’ sales on a critical day of the year, such as Christmas (McAfee & Brynjolfsson, 2012).

Variety: It encapsulates the increasingly different types of data, structured, semi-structured, and unstructured, from diverse data sources (*e.g.*, web, video and audio data, sensor data, financial data and transactional applications, log files and click streams, GPS signals from cell phones, social media feeds, and so on), and in different sizes from terabytes to zettabytes (IBM, 2016; Laney, 2001). One of the biggest challenges is posed by unstructured data. Unstructured data is a fundamental concept in Big Data and it refers to data that has no rules attached to it, such as a picture or a voice recording. The challenge is how to use advanced analytics to make sense of it.

Veracity: It refers to the trustworthiness of the data. In its 2012 Report, IBM showed that 1 in 3 business leaders don't trust the information they use to make decisions (IBM, 2012). One of the reasons for such phenomenon is that there are inherent discrepancies in the data, most of which emerge from the existence of unstructured data. This is even more interesting if we consider that, today, most of the data is unstructured. Another reason is the presence of inaccuracies. Inaccuracies can be due to the data being intrinsically inaccurate or from the data becoming inaccurate through processing errors (Laney, 2001).

Building upon the 4 Vs, Baesens *et al.* (2016, p. 807) stated that:

> *The 4V definition is but a starting point that outlines the perimeters. The definition does not help us to determine what to do inside the perimeters, how to innovatively investigate and analyze big data to enhance decision making quality, how to anticipate and leverage the transformational impacts of big data, or how best to consider scope as well as scale impacts of big data.*

As such, they argued for the necessity to include a 5^{th} V, namely *Value*, to complement the 4V framework. It should be noted that, by some accounts, there are as many as 10 Vs (Markus, 2015).

Charles and Gherman (2013) argued that the term Big Data is a misnomer, stating that while the term in itself refers to the large volume of data, Big Data is essentially about the phenomenon that we are trying to record and the hidden patterns and complexities of the data that we attempt to unpack. With this view, the authors advanced an expanded model of Big Data, wherein they included three additional dimensions, namely the 3 Cs: Context, Connectedness, and Complexity. The authors stated that understanding the Context is essential when dealing with Big Data, because "raw data could mean anything without a thorough understanding of the context that explains it" (p. 1072); Connectedness was defined as the ability to understand Big Data in its wider Context and within its ethical implications; and Complexity was defined from the perspective of having the skills to survive and thrive in the face of complex data, by means of being able to identify the key data and differentiate the information that truly has an impact on the organization.

Figure 3 briefly summarizes the discussed characteristics of Big Data. Presenting them all goes beyond the scope of the present chapter, but we hope to have provided a flavor of the various dimensions of Big Data. Having, thus, highlighted these, along with the existing debate surrounding the very definition of Big Data, we will now move towards presenting an overview of the social value that Big Data can offer.

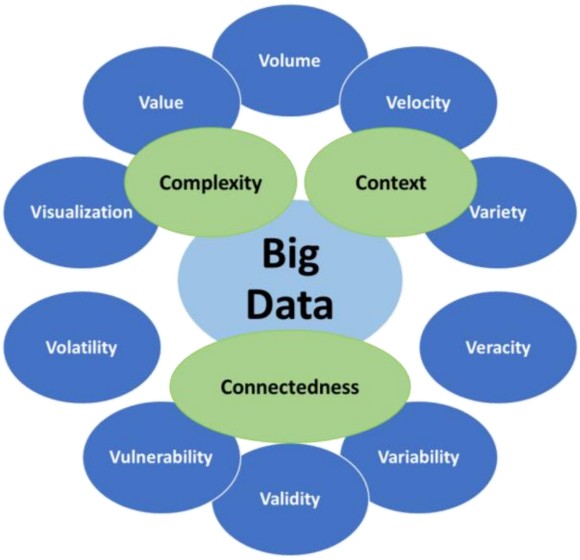

Figure 3. Big Data characteristics.

3 The Social Value of Big Data

Value creation in a Big Data perspective includes both the traditional economic dimension of value and the social dimension of value (Secundo, Del Vecchio, Dumay, & Passiante, 2017). Today, we are yet to fully understand how organizations actually translate the potential of Big Data into the said value (Günther, Rezazade Mehrizi, Huysman, & Feldberg, 2017). Generally, however, stories of Big Data's successes have tended to come from the private sector, and less is known about its impact on social organizations. Big Data can, nonetheless, drive big social change, in fields such as education (Cech, Spaulding, & Cazier, 2015; Pardos, 2017), healthcare (Raghupathi and Raghupathi, 2014), and public safety and security (Newell and Marabelli, 2015), just to mention a few.

Furthermore, the social value can be materialized as employment growth (Enck & Reynolds, 2009), increased productivity (Weill & Woerner, 2015), increased consumer surplus (Brynjolfsson & Saunders, 2010), new products and services, new markets and better marketing (Constantiou & Kallinikos, 2015), and so on. Governments, for instance, can use big data to, “enhance transparency, increase citizen engagement in public affairs, prevent fraud and crime, improve national security, and support the wellbeing of people through better education and healthcare” (Kim *et al*., 2014, p. 81).

Loebbecke and Picot (2015) published a viewpoint focusing on the impact of 'datification' (Galliers, Newell, Shanks, & Topi, 2015) on business models and employment and the consequential societal transformation. More recently, Günther *et al.* (2017) have conducted an in-depth systematic review of Information Systems literature on the topic and identified two socio-technical features of Big Data that influence value realization: portability and interconnectivity. The authors further argue that, in practice, "organizations need to continuously realign work practices, organizational models, and stakeholder interests in order to reap the benefits from big data."

In 2015, drawing upon Markus and Silver (2008) and Kallinikos (2011), Markus (2015, p. 58) elegantly drew attention to an important societal impact when she stated that the "source of Big Data's potentially transformative power lies not so much in the *characteristics* of Big Data, on which […] authors concentrate, as it does in what Big Data's characteristics *afford*." She continues by clarifying the statement: "Together with today's growing 'data lakes,' algorithms increasingly afford organizations the ability to automate the operational, and possibly even the strategic, decision making that is the core of managers' and knowledge workers' jobs".

As previously mentioned, the value that Big Data Analytics can unleash is great, but we are yet to fully understand the extent of the benefits. Empirical studies that consider the creation of value from Big Data Analytics are nowadays growing in number, but are still rather scarce. We, thus, join the call for further studies in the area. In the following lines, we will proceed to explore how Big Data can inform social change and to this aim, we present some of the advancements made in three different sectors. It should be noted that the information provided is not exhaustive, as our main intention is to provide a flavor of the opportunities brought about by the Big Data age.

Healthcare

Is it possible to predict whether a person will get some disease 24 hours before any symptoms appear? It is generally considered that the healthcare system is one of the sectors that will benefit the most from the existence of Big Data Analytics. Let us explore this in the following lines.

There is certain consensus that some of the challenges faced by the healthcare sector include the inadequate integration of the healthcare systems and the poor healthcare information management (Bodenheimer, 2005). The healthcare sector in general amasses a large amount of information, which nonetheless, results today in unnecessary increases in medical costs and time for both healthcare service providers and patients. Researchers and hospital managers alike are thus interested in how this information could be used instead to deliver a high-quality patient experience, while also improving organizational and financial performance and meeting future market needs (Agarwal, Gao, DesRoches, & Jha, 2010; Goh, Gao, & Agarwal, 2011; Ker, Wang, Hajli, Song, & Ker, 2014; Wang, Kung, Ting, & Byrd, 2015).

Watson (2014) and Raghupathi and Raghupathi (2014) advanced that Big Data Analytics can support evidence-based decision-making and action taking in healthcare. In this sense, Cortada, Gordon, and Lenihan (2012) performed a study and found that only 42% of the healthcare organizations surveyed supported their decision-making process with Big Data Analytics and only 16% actually had the necessary skills and experience to use Big Data Analytics. The value that can be generated in general goes, thus, far beyond the one that is created today.

But beyond improving profits and cutting down on operating costs, Big Data can help in other ways, such as curing disease or detecting and predicting epidemics. Big Data Analytics can help to collect and analyze the health data that is constantly being generated for faster responses to individual health problems; ultimately, for the betterment of the patient. It is now a well-known fact that with the help of Big Data Analytics, real-time datasets have been collected, modelled, and analyzed and this has helped speed up the development of new flu vaccines, identifying and containing the spread of viral outbreaks such as the Dengue fever or even Ebola.

Furthermore, we can only imagine for now what we would be able to do if, for example, we would collect all the data that is being created during every single doctor appointment. There are a variety of activities that happen during the routine medical examinations, which are not necessarily recorded, especially if the results turn out to be within some set parameters (in other words, if the patient turns out to be healthy): the doctor will take the body temperature and blood pressure, look into the eyes to see the retina lining, use an otoscope to look into the ears, listen to heartbeat for regularity, listen to breathing in the lungs, and so on – all these data could help understand as much about a patient as possible, and as early in his or her life as possible. This, in turn, could help identify warning signs of illness with time in advance, preventing further advancement of the disease, increasing the odds of success of the treatment, and ultimately, reducing the associated expenses. Now, to some extent, collecting this kind of granular data about an individual is possible due to smart phones, dedicated wearable devices, and specialized apps, which can collect data, for example, on how many steps a day a person walks and on the number of daily calories consumed, among others. But the higher value that these datasets hold is yet to be explored.

Psychiatry is a particular branch of medicine that could further benefit from the use of Big Data Analytics and research studies to address such matter have just recently started to emerge. It is well known that in psychiatric treatments, there are treatments that are proven to be successful, but what cannot be predicted generally is who they are going to work for; we cannot yet predict a patient's response to a specific treatment. What this means, in practical terms, is that most of the times a patient would have to go through various trials with various medicines, before identifying that works the best for the patient in question. In a recent contribution, Gillan and Whelan (2017) emphasized the importance of Big Data and robust statistical methodologies in treatment prediction research, and in so doing, they advocated for the use of machine-learning approaches beyond exploratory studies and toward model validation. The practical implications of such endeavors are rather obvious.

According to Wang, Kung, and Byrd (2018), the healthcare industry in general has not yet fully understood the potential benefits that could be gained from Big Data Analytics. The authors further stated that most of the potential value creation is still in its infancy, as predictive modelling and simulation techniques for analyzing healthcare data as a whole have not been yet developed. Today, one of the biggest challenges for healthcare organizations is represented by the missing support infrastructure needed for translating analytics-derived knowledge into action plans, a fact that is particularly true in the case of developing countries.

Agriculture

It is to some extent gratuitous to say that, in the end, nothing is more important than our food supply. Considering that we still live in a world in which there are people dying of starvation, it comes as quite a surprise to note that about a third of the food produced for human consumption is lost of wasted every year (Magnin, 2016). The agriculture sector is thus, in desperate need of solutions to tackle problems such as inefficiencies in planting, harvesting, and water use and trucking, among others (Sparapani, 2017). The Big Data age promises to help. For example, Big Data Analytics can help farmers simulate the impact of water, fertilizer, and pesticide, and engineer plants that will grow in harsh climatic conditions; it can help to reduce waste, increase and optimize production, speed up plant-growth, and minimize the use of scarce resources, such as water.

Generally speaking, Big Data Analytics has not yet been widely applied in agriculture (Kamilaris, Kartakoullis, & Prenafeta-Boldu, 2017). Nonetheless, there is increasing evidence of the use of digital technologies (Bastiaanssen, Molden, & Makin, 2000; Hashem *et al*., 2015) and bio-technologies (Rahman *et al*., 2013) to support agricultural practices. This is termed as *smart farming* (Tyagi, 2016), a concept that is closely related to *sustainable agriculture* (Senanayake, 1991).

Farmers have now started using high-technology devices to generate, record, and analyze data about soil and water conditions and weather forecast in order to extract insights that would assist them in refining their decision-making process. Some examples of tools being used in this regard include: agricultural drones (for fertilizing crops), satellites (for detecting changes in the field); and sensors on field (for collecting information about weather conditions, soil moisture and humidity, and so on).

As of now, Big Data Analytics in agriculture has resulted in a number of research studies in several areas – we herewith mention some of the most recent ones: crops (Waldhoff *et al.*, 2012), land (Barrett *et al*., 2014), remote sensing (Nativi *et al*., 2015), weather and climate change (Tesfaye *et al*., 2016), animals' research (Kempenaar *et al.,* 2016), and food availability and security (Frelat *et al*., 2016; Jóźwiaka, Milkovics, & Lakne, 2016).

According to Wolfert *et al.* (2017), the applicability of Big Data in agriculture faces a series of challenges, among which: data ownership and security and privacy issues, data quality, intelligent processing and analytics, sustainable integration of Big Data sources, and openness of platforms to speed up innovation and solution development. These challenges would need, thus, to be addressed in order to expand the scope of Big Data applications in agriculture and smart farming.

Transportation

Traffic congestion and parking unavailability are few examples of major sources of traffic inefficiency. Worldwide. But how about if we could change all that? How about if we could predict traffic jams hours before actually taking place and use such information to reach our destinations within lesser time? How about if we could be able to immediately find an available parking space and avoid frustration considerably? Transportation is another sector that can greatly benefit from Big Data. There is a huge amount of data that is being created, for example, from the sat nav installed in vehicles, as well as the embedded sensors in infrastructure.

But what has been achieved so far with Big Data Analytics in transportation? One example is the development of the ParkNet system ("ParkNet at Rutgers", n/a), a wireless sensing network developed in 2010 which detects and provides information regarding open parking spaces. The way it works is that a small sensor is attached to the car and an on-board computer collects the data which is uploaded to a central server and then processed to obtain the parking availability.

Another example is VTrack (Thiagarajan *et al*., 2009), a system for travel time estimation using sensor data collected by mobile phones that addresses two key challenges: reducing energy consumption and obtaining accurate travel time estimates. In the words of the authors themselves:

> *Real-time traffic information, either in the form of travel times or vehicle flow densities, can be used to alleviate congestion in a variety of ways: for example, by informing drivers of roads or intersections with large travel times ("hotspots"); by using travel time estimates in traffic-aware routing algorithms to find better paths with smaller expected time or smaller variance; by combining historic and real-time information to predict travel times in specific areas at particular times of day; by observing times on segments to improve operations (e.g., traffic light cycle control), plan infrastructure improvements, assess congestion pricing and tolling schemes, and so on.*

A third example is VibN (Miluzzo *et al*., 2011), a mobile sensing application capable of exploiting multiple sensor feeds to explore live points of interest of the drivers. Not only that, but it can also automatically determine a driver's personal points of interest.

Lastly, another example is the use of sensors embedded in the car that could be able to predict when the car would break down. A change in the sound being emitted by the engine or a change in the heat generated by certain parts of the car – all these data and much more could be used to predict the increased possibility of a car to break down and allow the driver to take the car to a mechanic prior to the car actually breaking down. And this is something that is possible with Big Data and Big Data Analytics and associated technologies.

To sum up, the Big Data age presents opportunities to use traffic data to not only solve a variety of existent problems, such as traffic congestion and equipment fault, but also predict traffic congestion and equipment fault before it actually happens. Big Data Analytics can, thus, be used for better route planning, traffic monitoring and management, and logistics, among others.

4 The Good… But What About The Bad?

As early as 1983 and as recent as 2014, Pool (1983) and Markus (2014), respectively, warned that Big Data is not all good. Any given technology is argued to have a *dual* nature, bringing both positive and negative effects that we should be aware of. Below we briefly present two of the latter effects.

In the context of Big Data and advanced analytics, a negative aspect, which also represents one of the most sensitive and worrisome issues, is the privacy of personal information. When security is breached, privacy may be compromised and loss of privacy can in turn result in other harms, such as identity theft and cyber-bullying or cyberstalking. "[…] There is a great public fear about the inappropriate use of personal data, particularly through the linking of data from multiple sources. Managing privacy is effectively both a technical and a sociological problem, and it must be addressed jointly from both perspectives to realize the promise of big data" (Ohlhorst, 2013, p. 122). Charles, Tavana, and Gherman (2015) advocated that, in the age of Big Data, there is a necessity to create new principles and regulations to cover the area of privacy of information, although who exactly should create these new principles and regulations is a rather sensitive question.

Marcus (2015) highlights another danger, which she calls 'displacement by automation'. She cites Frey and Osborne (2013) and notes that "Oxford economists […] computed (using machine learning techniques) that 47% of US employment is at risk to automation, though mainly at the lower end of the occupational knowledge and skill spectrum" (p. 58). We may wonder: *Is this good or is this bad?* On the one hand, behavioral economists would argue that humans are biased decision-makers, which would support the idea of automation. But on the other hand, what happens when individuals gain the skills necessary to use automation, but know very little about the underlying assumptions and knowledge domain that make automation possible?

It is without much doubt that The Bad or *dark* side of the Big Data age cannot be ignored and should not be treated with less importance than it merits, but more in-depth research is needed to explore and gain a full understanding of its negative implications and how these could be prevented, diminished, or corrected.

5 Conclusion

In this introductory chapter, we have aimed to provide an overview of the various dimensions and aspects of Big Data, while also exploring some of the key research studies that have been written and related applications that have been developed, with a special interest on the societal value generated.

It has been advocated that advanced analytics techniques should support, but not replace, human decision-making, common sense, and judgment (Silver, 1991; Charles & Gherman, 2013). We align with such assessment that indeed, without the qualities mentioned above, Big Data is, most likely, meaningless. But we also need to be pragmatic and accept evidence that points to the contrary. Markus (2015, p. 58), for example, pointed out that today there is evidence of automated decision-making, with minimal human intervention:

> *By the early 2000s, nearly 100% of all home mortgage underwriting decisions in the United States were automatically made (Markus et al.,*

> *2008). Today, over 60% of all securities trading is automated, and experts predict that within ten years even the writing of trading algorithms will be automated, because humans will not be fast enough to anticipate and react to arbitrage opportunities (The Government Office for Science, 2010). IBM is aggressively developing Watson, its Jeopardy game show-winning software, to diagnose and suggest treatment options for various diseases (Ungerleider, 2014). In these and other developments, algorithms and Big Data […] are tightly intertwined.*

It is, thus, not too bold to say that the potential applications and social implications brought by the Big Data age are far from being entirely understood and continuously change, even as we speak. Without becoming too philosophical about it, we would simply like to conclude the above by saying that a Big Data Age seems to require a Big Data Mind; and this is one of the greatest skills that a Big Data Scientist could possess.

Today, there are still many debates surrounding Big Data and one of the most prolific ones involves the questioning of the very existence of Big Data, with arguments in favor or against Big Data. But this is more than counter-productive. Big Data is here to say. In a way, the Big Data age can be compared to the transition from the Stone Age to the Iron Age: it is simply the next step in the evolution of the human civilization and it is, quite frankly, irreversible. And just because we argue over its presence does not mean it will disappear. The best we can do is accept its existence as the natural course of affairs and instead concentrate all our efforts to chisel its path forward in such a way so as to serve a Greater Good.

We conclude by re-stating that although much has been achieved until the present time, the challenge remains the insightful interpretation of the data and the usage of the knowledge obtained for the purposes of generating the most economic and social value (Charles, Tavana, & Gherman, 2015). We join the observation made by other researchers according to which many more research studies are needed to fully understand and fully unlock the societal value of Big Data. Until then, we will most likely continue to live in a world wherein individuals and organizations alike collect massive amounts of data with a ‘just in case we need it’ approach, trusting that one day, not too far away, we will come to crack the *Big Data Code*.

References

1. Agarwal, R., Gao, G., DesRoches, C., & Jha, A. K. (2010). Research commentary — the digital transformation of healthcare: current status and the road ahead. Information System Research, 21(4), 796-809.
2. Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Transformational Issues of Big Data and Analytics in Networked Business. MIS Quarterly, 40(4), 807-818.
3. Barrett, B., Nitze, I., Green, S., & Cawkwell, F. (2014). Assessment of multi-temporal, multisensory radar and ancillary spatial data for grasslands monitoring in Ireland using machine learning approaches. Remote Sensing of Environment, 152(2), 109-124.
4. Bastiaanssen, W., Molden, D., & Makin, I. (2000). Remote sensing for irrigated agriculture: examples from research and possible applications. Agricultural Water Management, 46(2), 137-155.
5. Beyer, M.A. and Laney, D. (2012) The Importance of 'Big Data': A Definition, META Group (now Gartner) [online] http://www.gartner.com/DisplayDocument?id=2057415 &ref=clientFriendlyUrl (accessed 1 August 2013).
6. Bodenheimer, T. (2005). High and rising health care costs. Part 1: seeking an explanation. Ann. Intern. Med. 142(10), 847-854.
7. Brynjolfsson, E. & Saunders, A. (2010). Wired for Innovation. How Information Technology is Reshaping the Economy. The MIT Press, Cambridge, MA.
8. Cech, T. G., Spaulding, T. K., & Cazier, J. A. (2015). Applying business analytic methods to improve organizational performance in the public school system. In: Proceedings of the Twenty-First Americas Conference on Information Systems, Puerto Rico, August, 13-15.
9. Charles, V. and Gherman, T. (2013). Achieving competitive advantage through big data. Strategic implications. Middle-East Journal of Scientific Research, 16(8), 1069-1074.
10. Charles, V., Tavana, M., & Gherman, T. (2015). The right to be forgotten – is privacy sold out in the big data age? International Journal of Society Systems Science, 7(4), 283-298.
11. Constantiou, I. D. & Kallinikos, J. (2015). New games, new rules: big data and the changing context of strategy. Journal of Information Technology, 30(1), pp. 44-57.
12. Cortada, J. W., Gordon, D., & Lenihan, B. (2012). The Value of Analytics in Healthcare: From Insights to Outcomes. IBM Global Business Services, Somers, NY.
13. Cox, M. & Ellsworth, D. (1997). Application-controlled demand paging for out-of-core visualization. Proceedings of the 8th IEEE Conference on Visualization. IEEE Computer Society Press, Los Alamitos, CA.
14. Davenport, T. H., Barth, P., & Bean, R. (2012). How 'big data' is different. MIT Sloan Management Review, 54(1), 43-46.
15. Einav, L. and Levin, J. (2013) 'The data revolution and economic analysis', Prepared for NBER Innovation Policy and the Economy Conference, April, [online] http://www.stanford.edu/ ~leinav/pubs/NBER2014.pdf (accessed 1 February 2014).
16. Enck, J. & Reynolds, T. (2009). Network Developments in Support of Innovation and User Needs, No. 164, OECD Publishing.
17. Fichman, R. G., Dos Santos, B. L., & Zheng, Z. (2014). Digital innovation as a fundamental and powerful concept in the information systems curriculum. MIS Quarterly, 38(2), 329-353.
18. Frelat, R., et al. (2016). Drivers of household food availability in sub-Saharan Africa based on big data from small farms. Proceedings of the National Academy of Sciences of the United States of America, 113(2), 458-463.

19. Frey, C. B. & Osborne, M. A. (2013). The Future of Employment: How susceptible are jobs to computerization? Oxford: Oxford Martin Programme on the Impacts of Future Technology.
20. Galliers, R., Newell, S., Shanks, G., & Topi, H. (2015). Call for papers for the special issue: the challenges and opportunities of 'datification'; Strategic impacts of 'big' (and 'small') and real time data – for society and for organizational decision makers. Journal of Strategic Information Systems, 24, II–III.
21. Gillan, C. M. & Whelan, R. (2017). What big data can do for treatment in psychiatry. Current Opinion in Behavioral Sciences, 18, 34-42.
22. Goh, J. M., Gao, G., & Agarwal, R. (2011). Evolving work routines: adaptive routinization of in- formation technology in healthcare. Information System Research, 22(3), 565-585.
23. Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. Journal of Strategic Information Systems, 26, 191-209.
24. Hammond, K.J. (2013) 'The value of big data isn't the data', Harvard Business Review, May [online] http://blogs.hbr.org/cs/2013/05/the_value_of_big_data_isnt_the.html (accessed 13 July 2013).
25. Hashem, I., et al. (2015). The rise of "big data" on cloud computing: review and open research issues. Information Systems, 47, 98-115.
26. Herschel, R., & Miori, V. M. (2017). Ethics and Big Data. Technology in Society, 49, 31-36.
27. IBM. (2012). Simply Good Design. 2012 IBM SOA Architect Summit. Retrieved July, 20, https://www950.ibm.com/events/wwe/grp/grp004.nsf/vLookupPDFs/SOA%20Design%20Principles%20and%20Development%20Q4%202012_Final/$file/SOA%20Design%20Principles%20and%20Develop ment%20Q4%202012_Final.pdf.
28. IBM. (2013). What is big data? Retrieved 20 July 2012, from http://www-01.ibm.com/software/data/bigdata.
29. IBM. (2016). What is Big Data? Retrieved 20 November 2017, from https://www-01.ibm.com/software/data/ bigdata/what-is-big-data.html
30. Jóźwiaka, Á., Milkovics, M., & Lakne, Z. (2016). A network-science support system for food chain safety: a case from Hungarian cattle production. The International Food and Agribusiness Management Review, Special Issue, 19(A).
31. Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldu, F. X. (2017). A review on the practice of big data analysis in agriculture. Computers and Electronis in Agriculture, 143, 23-37.
32. Kempenaar, C., et al. (2016). Big Data Analysis for Smart Farming, vol. 655 Wageningen University & Research, s.l.
33. Ker, J. I., Wang, Y., Hajli, M. N., Song, J., & Ker, C. W. (2014). Deploying lean in healthcare: evaluating information technology effectiveness in US hospital pharmacies. International Journal of Information Management, 34(4), 556-560.
34. Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. Applications Delivery Strategies, META Group (now Gartner) [online] http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling- Data-Volume-Velocity-and-Variety.pdf (accessed 1 August 2013).
35. Loebbecke, C. & Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: a research agenda. The Journal of Strategic Information Systems, 24(3), 149-157. http://dx.doi.org/10.1016/j.jsis.2015.08.002.
36. Magnin, C. (2016). How big data will revolutionize the global food chain. [online] McKinsey & Company. Retrieved from https://www.mckinsey.com/business-

functions/digital-mckinsey/our-insights/how-big-data-will-revolutionize-the-global-food-chain

37. Markus, L. (2015). New games, new rules, new scoreboards: the potential consequences of big data. Journal of Information Technology, 30(1), 58-59.
38. Markus, M. L., Dutta, A., Steinfield, C. W., & Wigand., R. T. (2008). The Computerization Movement in the Us Home Mortgage Industry: Automated underwriting from 1980 to 2004. In K. L. Kraemer and M. S. Elliott (Eds.), Computerization Movements and Technology Diffusion: From mainframes to ubiquitous computing (pp. 115-144), Medford, NY: Information Today.
39. McAfee, A. & Brynjolfsson, E. (2012). Big data: the management revolution. Harvard Business Review, 90(10), 60-68.
40. McKinsey Global Institute (2013) Game Changers: Five Opportunities for US Growth and Renewal, July [online] http://www.mckinsey.com/insights/americas/us_game_changers (accessed 13 December 2013).
41. Miluzzo, E., Papandrea, M., Lane, N. D., Sarroff, A. M., Giordano, S., & Campbell, A. T. (2011). Tapping into the vibe of the city using vibn, a continuous sensing application for smartphones. In Proceedings of 1st international symposium on From digital footprints to social and community intelligence (pp. 13-18). Beijing, China: ACM.
42. Nativi, S. et al. (2015). Big data challenges in building the global earth observation system of systems. Environmental Modelling & Software, 68(1), 1-26.
43. Newell, S. & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: a call for action on the long-term societal effects of 'datafication'. The Journal of Strategic Information Systems, 24(1), 3-14. http://dx.doi.org/10.1016/j.jsis.2015.02.001.
44. Ohlhorst, F. (2013) Big Data Analytics: Turning Big Data into Big Money, John Wiley & Sons Inc, Hoboken, NJ.
45. Pardos, Z. A. (2017). Big data in education and the models that love them. Current Opinion in Behavioral Sciences, 18, 107-113.
46. Raghupathi, W. & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health Information Science and Systems, 2(3), 1-10. http://dx.doi.org/10.1186/2047-2501-2-3.
47. Rahman, H., Sudheer Pamidimarri, D., Valarmathi, R., & Raveendran, M. (2013). Omics: Applications in Biomedical, Agriculture and Environmental Sciences, s.l. CRC PressI Llc.
48. Secundo, G., Del Vecchio, P., Dumay, J., & Passiante, G. (2017). Intellectual capital in the age of Big Data: Establishing a research agenda. Journal of Intellectual Capital, 18(2), 242-261.
49. Senanayake, R. (1991). Sustainable Agriculture: definitions and parameters for measurement. Journal of Sustainable Agriculture, 1(4), 7-28.
50. Sparapani, T. (2017). How Big Data and Tech Will Improve Agriculture, from Farm to Table. [online] Forbes. Retrieved from https://www.forbes.com/sites/timsparapani/2017/03/23/how-big-data-and-tech-will-improve-agriculture-from-farm-to-table/#503f16c25989
51. Tesfaye, K., et al. (2016). Targeting drought-tolerant maize varieties in southern Africa: a geospatial crop modeling approach using big data. International Food and Agribusiness Management Review, 19(A), 1-18.
52. The Government Office for Science. (2010). Foresight: The Future of Computer Trading in Financial Markets, Final Project Report, London.
53. Thiagarajan, A., Ravindranath, L., LaCurts, K., Madden, S., Balakrishnan, H., Toledo, S., & Eriksson, J. (2009). Vtrack: accurate, energy-aware road traffic delay estimation using mobile phones. In Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (pp. 85-98). Berkeley, California: ACM.

54. Tyagi, A. C. (2016). Towards a second green revolution. Irrigation Drainage, 65(4), 388-389.
55. Ungerleider, N. (2014). IBM's Watson is ready to see you now – in your dermatologist's office. Fast Company: May 16, [www document] http://www.fastcompany.com/3030723/ibms-watson-is-ready-to-see-you-now-in-yourdermatologists-office
56. Waldhoff, G., Curdt, C., Hoffmeister, D., & Bareth, G. (2012). Analysis of multitemporal and multisensor remote sensing data for crop rotation mapping. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 25(1), 177-182.
57. Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting & Social Change, 126, 3-13.
58. Wang, Y., Kung, L., Ting, C., Byrd, T. A. (2015). Beyond a technical perspective: understanding big data capabilities in health care. System Sciences (HICSS), 2015 48th Hawaii International Conference (pp. 3044–3053). IEEE.
59. Watson, H. J. (2014). Tutorial: big data analytics: concepts, technologies, and applications. Communications of the Association for Information Systems, 34(1), 1247-1268.
60. Weill, P. & Woerner, S. (2015). Thriving in an increasingly digital ecosystem. MIT Sloan Management Review, 56(4), 27-34.
61. Why Walmart Always Stocks Up On Strawberry Pop-Tarts Before a Hurricane. (2017). Retrieved from http://www.countryliving.com/food-drinks/a44550/walmart-strawberry-pop-tarts-before-hurricane/
62. Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big Data in Smart Farming – A Review. Agricultural Systems, 153, 69-80.