



S. Klous
is a senior manager at
KPMG Advisory.
klous.sander@kpmg.nl

Big Data: Big Opportunities, Big Responsibilities

Sander Klous

Big Data have often been referred to as the “New Oil”, thereby raising expectations of huge economic potential. New business models are emerging in all sectors based on the smart use of data, such as the travel industry, the music industry and retail. However, there’s much more to Big Data than commercial opportunities. In a broader sense, Big Data is an important formative force within a new information society: It brings new concepts to improve our lives and to better organize processes and other forms of societal value. Inevitably, the rise of Big Data also implies new risks and dangers, such as in the privacy domain. Establishing equilibrium between profiting from Big Data and managing personal concerns is an issue we have to resolve.

Introduction

Big Data has often been referred to as the “New Oil”. Many organizations have just started to unlock this potential and hope to profit from the new opportunities that smart use of data can offer. We see radical changes taking place all around us: new business models are emerging in all sectors based on the smart use of data. However, Big Data covers much more than purely commercial prospects. Big Data is becoming an important shaping factor in a new information society, bringing new concepts to improve our lives, to organize complex processes and other forms of societal structure. Of course, Big Data also nourishes the development of new risks and dangers, as in the privacy domain for example. The challenge is to enjoy the positive side of Big Data while effectively controlling these new risks.

It’s the Numbers That Count

New technology has made incredible things possible. And new technology is now part of our lives in a way we could scarcely have dreamed of 20 years ago. A new information society is appearing, a world in which everything is

measurable and in which people – and almost every device you can think of – are connected through the internet on a 24/7 basis. That network of connections and sensors provides a phenomenal amount of data, and offers fascinating new possibilities which are often collectively called “Big Data” (see box for some thoughts on the definition of Big Data). One of the radical changes is: we no longer have to ask people what they think or feel, we simply measure what they do. This brings a vast amount of opportunities for tailor-made services. In this article, we will not explore the technical aspects of Big Data, but rather focus on the consequences related to bringing new value.

*“Moore’s law tells us
that data are becoming
exponentially incomputable”*

– Cees de Laat, University of Amsterdam

What Is Big Data?

Big Data is often described in terms of 5 Vs: Volume, Variety, Velocity, Veracity and Value.

Volume refers to the vast amounts of data that are being generated. The Internet of Things is linking up billions of devices – TVs, refrigerators, security devices, thermostats, smoke detectors – all of which produce and share data. This contributes to the growth of the volume of data. Many data sets have simply become too large to store, process and analyze by means of traditional database technology. An interesting fact in this respect: Moore's law tells us that data are becoming exponentially incomputable: storage capacity is doubling every 9 months, but CPU capacity is only doubling every 18 months. The only way to address this challenge is to become smarter about which data to analyze, or to create smarter algorithms.

Variety refers to Big Data addressing challenges beyond the analysis of data with clear structures, such as financial figures: there is great variety in the nature of the data. All our chatter on social media, streaming music and videos, our e-mails, our online shopping transactions and so on have no clear pre-determined structure. Many new tools are largely generic, allowing collection and analysis of different types of data, such as messages, social media conversations, photos, sensor data, video and voice recordings, etc. Combining insights in these types of data with classical data analysis can result in a more profound understanding of the behavior of (groups of) people, systems, or processes.

Velocity refers to both the increasing rate of data collection and the increasing demand to acquire real-time insights and responses. These systems come with their own challenges. On the collection side, decisions are made concerning which data are to be kept and which are to be discarded. Wrong selections are irreversible. Systematic errors are harder to detect when only a biased sample of the data is available. On the decision side, there are several examples of cases in which data analysis has been taken out of context, or has been simply wrong, resulting in serious instabilities in an ecosystem. The most famous example is probably the flash crash on Wall Street on 6 May 2010. The Dow Jones index lost 10% of its value within minutes, only to recover most of its value shortly afterwards. The crash was attributed to faulty automatic trading algorithms, making real-time decisions without human intervention.

Veracity refers to the correctness and accuracy of information. In the era of Big Data, perception of this topic is changing. In the past decade, masterdata management, data quality and data governance were at the heart of most of the endeavors of large organizations to get insight in their data. Although the quality of data remains an important topic, the pervasiveness of data makes it impossible to manage and control the quality of all sources. This is often already true within larger organizations, but becomes especially relevant when the data of other organizations are included in an analysis. Still, it is important to understand the reliability of datasets. This can often be achieved by processing an increasing amount of data, following the “quantity over quality” mantra, either by increasing the sample size to reduce statistical errors or by adding independent sources to reduce systematic errors. Although this might sound like a straightforward solution, advanced statistics and modeling techniques are often required to be able to estimate properly the impact of combining datasets.

The first four Vs focus on technical elements of Big Data. However, one of the most distinguishing features of Big Data, compared to other data-related subjects, is the fact that it is primarily a topic driven by business, not by technology. Hence, arguably the most interesting “V” of the five, is the **Value** it can bring. In the main text, this is one of the central story lines.

Despite the 5V definition, we use Big Data as an overarching term in this article. It covers all the new opportunities, possibilities, threats and techniques associated with the fact that we can deal with data in a different way, such as the positive and negative sides of the “datafication” of society for example, including social elements such as privacy and social influences.

Other Side of the Coin

To examine Big Data properly, it is important to realize that it is more than just a new technology. Big Data offers opportunities to all sectors and industries to organize ourselves differently, to make progress, and to do things that were simply not possible until recently. Many people associate the term Big Data with companies wanting to sell their customers more stuff by learning everything there is to know about them. But that's only one side of the coin: Big Data can also solve social problems and improve our lives. We can use it to save human lives. We can improve maintenance and plan more cost-effectively. We can improve traffic safety. We can increase agricultural yields. We can provide a better quality of life for the elderly. We can make the world more sustainable. We can revolutionize medical diagnostics and treatment. We can increase the security of payments. We can improve a team's sporting performance. And we can reduce traffic congestion.

Public and private organizations are starting to realize this. They are asking themselves how they ought to respond to these new opportunities. Clearly, this goes beyond launching new products or services: it's about how to respond to a totally different world. In this respect, the term "age of disruption" is increasingly being used. The term describes the disruptive effect that rapid successive technological breakthroughs are having on virtually all parts of society.

Fail or Scale

The speed of change is phenomenal in certain technological areas, bringing an extra challenge to many companies. Developing and successfully deploying a groundbreaking new technological concept may be fine, but it is no reason to sit back and enjoy life. In today's "flat" world, there is a continuous danger that a greenfield competitor will pop up with an even better concept. This new concept can then quickly gain ground and be scaled. A vivid example of this type of development is the way Netflix, the streaming company, is conquering large parts of the world, racing ahead of traditional entertainment. It is therefore a matter of fail or scale for many companies that are active in the field of Big Data and/ or the Internet of Things.

In other words, Creative Destruction is occurring at the speed of light. The core of this economic theory is that we can only achieve innovations when the old models are being torn down. This is happening all around us at the

moment. The entertainment industry has already had its share of these disruptive forces. More recently, Airbnb brought radical changes to the travel industry and Uber is currently trying to shake up the taxi industry. These companies approach the application of data in a totally different way than conventional companies do in their domain. In fact, Uber is not a taxi business. It's in the business of selling customer data to a network of taxi drivers and of optimizing the taxi network by bringing in more data intelligence. The customer is actually the product. Other sectors will probably follow suit and will also have to deal with newcomers with disruptive new models based on the use of (Big) Data.

Ethical Discussions

In the meantime, ethical discussions about the uses of Big Data are cropping up on an increasing scale. To many of us, Big Data has negative connotations. People are wary that companies or governments may be abusing personal information and/ or feel that surveillance of their personal activities is extending beyond their level of comfort.

There is not much discussion on the issue of privacy in cases where organizations are trying to optimize operational efficiency with predictive maintenance for example. But in other cases, the general public can be very sensitive and critical when it comes to the issue of how organizations can bring benefit to our well-being, or even to society, with their Big Data solutions. The healthcare sector is an evident example of this: the use of personal data from various sources may bring new and better forms of medical care – based on tailor-made approaches rather than a one-size-fits-all approach – but is also a very sensitive domain in terms of the use of personal data. As a consequence, organizations must go beyond their obligation to simply obey the law in order to obtain a license to operate. If they fail to show that value for their customers is a central goal, they will be perceived as cherishing evil intentions and they thus put their brand value and reputation at stake. The challenge for companies is to build data-driven strategies that offer a combination of value to the customer or society as a whole – in terms of comfort, safety, etc. – as well as commercial value. Here, we refer to Michael Porter's ideas on Shared Value: this combination of commercial and societal value should be central to every data-driven strategy. In practice, however, many companies struggle to explore this combination, focusing solely on the commercial opportunities of data, thereby scaring off their customers.

Revolutionize Healthcare

It is essential to find ways to deal properly with the (privacy) dilemmas that arise with the use of Big Data, as the stakes are high. As mentioned earlier, healthcare may serve as an example of this, as Big Data can truly stimulate a quantum leap in this domain. The chances are that it is only a matter of time before we transform cancer from a killer disease into a chronic disease. The key to this is mapping and analyzing DNA. The shift in this field is profound. By 2001, scientists were able to unravel a full strand of human DNA after pouring hundreds of millions into research and computing power. Nowadays, commercial companies offer this service for a few hundred dollars. The combination of personal medical data – such as DNA profiles – with lifestyle and other parameters will offer unparalleled opportunities for tailor-made prevention and diagnosis. Worldwide, hundreds of companies are now jumping on the bandwagon to unlock this potential.

Their efforts can only be sustainably successful if we find a way to address the concerns accompanying Big Data, in order to ensure that we can profit from the societal progress it can bring. This is far from new: throughout history, new technology has always had a bright and a dark side. Once mankind developed the stone axe, we were not only able to make better instruments, we could also use this stone axe to kill our neighbor.

Ecosystem

The main challenge is to build an ecosystem in which we can anchor fundamental principles such as privacy, transparency, and the right to be forgotten. Technologically, we can definitely do this. Several techniques are available to use personal data in such a way that privacy is not compromised (pseudonymization and aggregation).

We believe that this ecosystem will be built in the years to come and several initiatives and pilot projects also point into this direction. Technology can be used as an instrument for controlling the negative sides in this ecosystem. To understand this, we must realize that software and hardware will increasingly determine our behavior and probably much more so than laws and regulations. Google's algorithm determines which items are relevant to you: the software and hardware of the smartcard system determines how you use public transport. "The system" is becoming increasingly dominant in determining our behavior.

To summarize this development, we could paraphrase Lawrence Lessig by stating that: "Code is law". That in itself is no bad thing, because we can change the code, meaning our systems, to our liking if we want to. However, we need to make an important point. When a system becomes the equivalent of the law, the system developer is the legislator. Therefore it is essential that the architecture of the systems we use is aligned with what we want as a society. And this is where things often go wrong. The principles that we adopt in society – in the areas of privacy, transparency, "the right to be forgotten" for example – are not the first priority in the tasks we assign to system developers, whereas they should be. If they were, our privacy rights, for example, would be much better safeguarded. In their construction, the setup of systems must be aligned with the principles underlying existing legislation. Technology can then be used to ensure our fundamental rights, including privacy.

“We must realize that software and hardware are increasingly determining our behavior and probably much more so than laws and regulations”

– derived from the famous phrase “Code is Law” of Lawrence Lessig, Stanford



Conclusion

To conclude, we wish to emphasize that we need to be careful not to identify Big Data with number crunching, as the process of data analysis in itself should be in the hands of trained professionals. We refer to one of the many Sherlock Holmes quotes, which states that: “It’s human nature to see only what we expect to see”.

This should be a key warning to anyone involved in data analysis. We should prevent accidents occurring as a result of improper interpretation of data. This is a real risk. When gathering massive amounts of data and performing various analyses, it is inevitable that a striking correlation will pop up. This may not be a very relevant correlation, however, or may even be a dangerous conclusion. One example is the correlation between butter production in Bangladesh and the development of the S&P 500 stock index. Obviously, this is not a very useful finding. Suppose that the previously mentioned data revolution in healthcare gains traction, we will need trained data scientists to look at the data. After all, we prefer to be treated on the basis of causality, but might need to settle for a poorly understood correlation. The new ecosystem will only function well in the hands of skilled data scientists. A fool with a tool is, after all, still a fool.

References

[Klou14] S. Klous and N. Wielaard, *Wij zijn Big Data* (We Are Big Data), Business Contact, 2014.

About the Author

S. Klous is a professor in Big Data Ecosystems at the University of Amsterdam, and Managing Director Big Data Analytics at KPMG Advisory. He is head of technology in the KPMG global leadership team on Data & Analytics. His book *Wij zijn Big Data* (“We Are Big Data”), about the future of the information society, was published in 2014 and was nominated for the management book of the year award 2015. An updated edition will be published in English by Springer in the fall of 2015.

*“It’s human nature
to see only what we
expect to see”*

– Sherlock Holmes