

**ESADE**

Ramon Llull University

ESADEgeo-CENTER  
FOR GLOBAL ECONOMY  
AND GEOPOLITICS

**E**

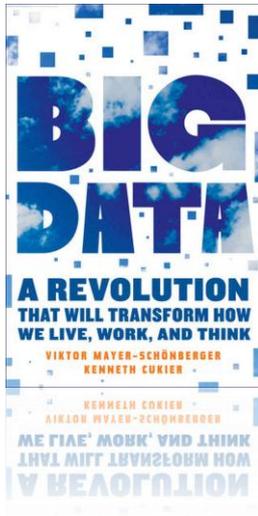
**30**

**Book reviews  
on global economy  
and geopolitical  
readings**

*ESADEgeo, under the supervision of Professor Javier Solana  
and Professor Javier Santiso.*



# Big Data: A Revolution that Will Transform How We Work, Live and Think



**Mayer-Schönberger, Viktor y Cukier, Kenneth, (2013), Houghton Mifflin Harcourt, Boston, New York.**

*“Big data is a resource and a tool. It is meant to inform, rather than explain; it points us towards understanding; but it can still lead to misunderstanding, depending on how well or poorly it is wielded.”*

*“Today data refers to a description of something that allows it to be recorded, analysed and reorganized.... Let’s call it datafication.”*

*“Big data refers to things that one can do at large scale that cannot be done at smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments and more.”*

## Summary

Big Data enabled Google to predict the spread of the H1N1 influenza virus, helps the police identify the most dangerous New York sewers before they explode, and gives Amazon the ability to recommend books that really interest its clients. This term ‘Big Data’ describes the newly acquired ability to collect a huge amount of information, instantly analyse it, and draw diverse conclusions for various applications.

Obviously, the rapid development of technological tools in recent years has been key to the flowering of this new activity; however, the authors, Viktor Mayer-Schönberger and Kenneth Cukier, confirm that the decisive factor has been the increasing use of mass data collection. Everything entered into social networks or search engines, for example, is collected and stored. This approach means that information is datafied so that computer applications can analyse and extract conclusions we could never have imagined.

We are currently witnessing a revolution in the dissemination of new information that will be compared with the invention of the printing press, or the birth of the internet. Big Data changes our perception of business, politics, health, or the economy; yet, like all revolutions, it also involves serious risks that must be taken into account. The end of privacy, the danger of its predictive power, or the threat that information will dominate us rather than the other way round, are part of the dark side of Big Data. To

avoid the dangers, the authors say we need a new legal system and specific monitoring mechanisms.

## The authors

**Viktor Mayer-Schönberger** is professor of governance and regulation at the Oxford Internet Institute, University of Oxford, where he arrived after ten years at John F. Kennedy School of Governance at Harvard. He has written eight books and over a hundred articles on Big Data. His most recent solo volume was *Delete: The Virtue of Forgetting in the Digital Age*. He also works as a consultant to major corporations and international organisations such as Microsoft and the World Economic Forum.

**Kenneth Cukier** is data editor of *The Economist* and an expert on Big Data. He previously worked as a correspondent for *The Economist* in Tokyo, *The Wall Street Journal Asia* in Hong Kong, and the *International Herald Tribune* in Paris. His articles on economics and business have been published in *Foreign Affairs*, *The New York Times* and the *Financial Times*.

## Key ideas and opinion

This book is the first theoretical approach to a new reality, the world of Big Data, which affects not only internet users as individuals, but sectors such as healthcare, business, or politics. **Every little piece of data generated by a person, community, or a country is collected and saved.** This has created a huge archive of information, which thanks to technological advances, can be chewed over by mathematical algorithms and then analysed. Amazing and useful conclusions can be drawn in a variety of fields from this vast heap of information. Viktor Mayer-Schönberger and Kenneth Cukier insist that the importance of this change is not so much the tools used, but Big Data itself, namely, the huge volume of information generated by internet searches, social networks, and geolocation programs.

The authors divide their book into three parts. The first presents and **describes the term Big Data**: what it is, how it is developed, its strengths, and how the concepts of error and causation have given way to disorder and correlation. The sheer volume of data means that we are satisfied with mere correlation – the ‘why’ or the causality is no longer needed. The book then goes on to analyse the concepts of correlation and datafication. The second part discusses **the business value** of this information and the emerging data market. Finally, the book presents **Big Data’s dark side**: the main dangers being the violation of privacy, and prejudice and tyranny of information. Big Data will change our perception of reality as did the printing press or the internet, and we must prepare to protect ourselves. This is **an interesting and useful book for**

# E

understanding a reality in which every individual is seen as a generator of valuable information.

## Big Data

In 2009, the new **H1N1** flu virus kept American health services and politicians on tenterhooks because they were unable to monitor its spread. **Google** saved the day by creating a system that sought correlations between terms that users keyed into the search engine and the spread of the flu. This showed that a new method, more effective than traditional statistical studies, could be used to achieve real-time solutions. **It was Big Data: society's ability of to use information in new ways to produce useful knowledge and services with significant value.** Google's case is just one example of the great changes introduced by this technology. However, the authors claim that the change in mindset about how billions of pieces of data can be used is even more important. Say the authors: "Big data refers to things that one can do at large scale that cannot be done at smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments and more." In this new vision, the search for causality becomes a search for correlation: expecting to find the 'what', not the 'why'.

Google processes over 24 petabytes of data a day, Facebook receives more than 10 million photos daily, and Twitter is growing by 200% annually with 400 million tweets a day. **The volume of information is growing rapidly** and Big Data has been positioned as the system for applying mathematics to this huge volume of information and then calculating probabilities. The authors point to **three interconnected and essential changes of mentality**: the ability to analyse a vast amount of information on a topic; the willingness to accept somewhat messy data about the real world instead of looking for accuracy; and finally, a growth of respect for correlation rather than an insistence on causation.

The authors explain this **disorder** in a simple way: **the more data you have, the more errors and inaccuracies you will find.** However, there is no need to reduce the errors. **This degree of chaos is an acceptable trade-off for the larger amount of information,** which, in turn, creates more precise conclusions. As information is no longer scarce (quite the contrary), an obsession with accuracy is obsolete, and as a society, we must learn to see the world from this new and more complete perspective.

An important concept coined by the authors to define this activity is **datification: transforming a phenomenon into a format that can be quantified for tabulation and analysis.** It should be clear that the act of digitisation does not in itself lead to datification. This difference is clearly seen in the digitisation of books by Google. By itself, the information in these books cannot be used for more than reading. Datification means the data is made accessible to computers, which can analyse it through algorithms, and the text becomes searchable in what is termed **culturomics**: a

form of computational lexicology that can be used to understand human behaviour and cultural trends through the quantitative analysis of texts.

Geolocation is another major generator of information and the ability to collect the location of users has become very useful, for example, for **customising advertising**. These indirect (commercial) uses are quite different from the initial use that was intended for these systems. Datafication has created new uses and values **in aspects of our personal lives** that can be applied to relationships, experiences, and even moods. Twitter, for instance, sells your information to two companies, and DataSift and Gnip, for what has been called '**sentiment analysis**'. This datafication leads, among other things, to forecasts of box office receipts for films, and stock market price forecasts for shares. All thanks to an analysis of the moods behind tweets.

## Value

Private companies are currently the most advanced datafication users, and information has become an attractive source of income for many firms, which unlike material goods, does not depreciate when used and can be repeatedly processed. The authors have summarised three emerging **strategies to increase the value of the data**. Firstly, data reuse. An example is the storage of search terms so that they can then be combined to establish user preferences. The book points to the case of BBVA, which has joined with Google to analyse the tourism industry and sell real-time economic indicators based on search results. Even the most ordinary information has value if applied correctly, and there are many companies who can see an industry emerging from these techniques. Telefónica, for example, has created a separate company called Telefónica Dynamic Insights to sell stores and other clients anonymous information about the location of its users.

Another way to create value is by **combining different datasets with application options**. This is the case of Google Street View, whose famous cars collect far more data than is needed for the main function of the program, with the data then being applied for secondary purposes, such as improving the map service. The cost of collecting all the data possible is not very great, and so it seems logical to do so, with a view to future projects. **This is what the authors call 'twofers': when a dataset can be used for several applications (if collected in a specific way) and so double its usefulness.**

**Although the value of data seems clearly proven, it is difficult to set a price** in the era of Big Data. It is not enough to account for its first use, because greater – and initially, incalculable – value often lies in subsequent developments. The **owners of data often prefer a system similar to the percentages that publishers pay book authors**. That is, they charge a portion of the value extracted rather than a flat rate.

A busy and expanding data market has been established and many companies are specialising in the trade, such as Decide.com, which scans data for opportunities, discovering the jewels beneath the masses of records. There are **three types of companies working in this market**:

1. Those that **hold the data**: for example banks, which can analyse the information themselves, sell it, or contract others to analyse it for them.
2. Those with the **ability to analyse data**: specialised companies with the right technologies.
3. Those with a Big Data mentality, who have **unique ideas on how to take advantage of data** and increase its value.

Firms from the first two categories are currently capturing most of the attention, but the third category is gradually gaining ground with innovative ideas. However, the authors believe that information will be revalued in the future and those with the data will dictate the terms. Proof of this trend is the \$700 million that Google paid in 2008 for the data provider ITA Software.

### The dark side

Like any new activity, Big Data involves new risks. A new legal framework is needed to protect the public.

**The first problem is privacy**, which is seeing its boundaries pushed aside in favour of extended data collection. Much of the seemingly limitless information that is recorded is personal. With Big Data, the value of this data is not only in the first use, but in the subsequent uses that follow – and so the classic notice ‘I have read and accept the conditions’ is no longer valid. **The user only gives permission for the first use made of the information, but not the next, and as of yet, unknown use.** Neither do the other two control mechanisms, the exemption clause and anonymity, offer much protection, as these controls are likely to be forgotten along the way (a good example is the use the NSA made of data obtained through search engines).

The authors propose the creation of new systems to safeguard privacy that are less focused on user consent and more focussed on the responsibility of the owners of such information: **legislation is needed on how to assess the usage risk and determine how to prevent potential damage.** Violation of these new rules would be a criminal offense – but for the company, compliance will offer the advantage of not having to consult the original data generator. **Shifting the burden of responsibility from the public to those who use the information** is thus the first step, which can then be reinforced with a **time limit on use** – after which data must be discarded.

The second risk is based on concepts of **probability and punishment** – and is perhaps the most dangerous. Big Data creates predictions and may define, for example, the likelihood that an individual may commit a crime. Preventive policies already use Big

# E

Data to establish which individuals, neighbourhoods, or cities should be subject to increased surveillance to prevent crime. However, this creates a stigma, and perhaps even the desire to punish the possible criminal without any crime being committed. Such judgments, the authors insist, destroy the presumption of innocence and the ability of human beings to make moral choices. The authors propose as control measures: transparency, certification, and the possibility of rejecting data about oneself. **There must be monitoring performed by people to ensure that individuals are judged on their deeds, and not by what an algorithm predicts they might do.**

Thirdly, there is **a risk of converting data into an absolute truth and creating a data dictatorship**. Here the authors recall the experience of Robert McNamara, American secretary of defense under Kennedy, whose faith in statistics had fatal consequences during the Vietnam War: an example of the limits of information that should be remembered. Monitoring and transparency are the emerging solutions – and these will need new institutions and experts to carry them out. A new type of professional is needed for this function – the *algorithmist*. Such professionals may be independent outsiders or part of a corporate department. **As experts in computing, mathematics, and statistics, they will strive to guarantee confidentiality and impartiality with a Hippocratic oath similar to doctors and lawyers.** Their work should focus on monitoring current Big Data practices. Finally, antitrust legislation should protect us as much as possible from a data market that is fed daily with vast volumes of valuable personal data and whose rapid growth means that control mechanisms should be as effective as they are instant.