LUISS Guido Carli
LIBERA UNIVERSITÀ INTERNAZIONALE DEGLI STUDI SOCIALI

DEPARTMENT OF ECONOMICS AND FINANCE

CHAIR OF LAW AND ECONOMICS

EVER BIGGER DATA: HOW ANALYTICS ARE BECOMING A GROWING PARADIGM

Thesis Supervisor:

Prof. Andrea Renda

Candidate:

Laura Garofalo

160061

Academic Year 2013-2014

To my parents.

Contents

ABSTRACT

This thesis analyses how Big Data revolutionized the world we are living in.

The basic question this study will try to answer to is: Big Data is affecting our lives in a positive way or not? We will analyse all the possible ways in which Big Data can be employed and define the pros and cons to then end up with, hopefully, a clear idea of whether to trust this revolution or not.

In the first chapter we will have a brief introduction to the world of Big Data; while in the second chapter we will discover its' origins, present and some future developments. In chapter three the study will focus on analysing how Big Data in employed in the different sectors such as healthcare, crime prediction and marketing. Finally, in chapter four the study also explains the problems arising from all the privacy debates that there are regarding the use of personal data.

CHAPTER 1

1.1 Towards a Big Data society

"Big data is the derivation of value from traditional relational database-driven business decision making, augmented with new sources of unstructured data." (Oracle)

"Big data is the term increasingly used to describe the process of applying serious computing power—the latest in machine learning and artificial intelligence—to seriously massive and often highly complex sets of information." (Microsoft)

It is hard to give a definition to something so elaborate and debatable but we could say that we ourselves are Big Data or at least, we create it.

Our life is surrounded by Big Data and most of the times we unknowingly use it. We wake up with the custom sound of our cell phone which immediately reminds us of our meetings, birthdays and shopping lists then we go to our kitchen to have breakfast and our technological fridge which tells us that we are running out of eggs; we get in our car to go to work and our GPS turns on to warns us about traffic alerts caused by and accident; we Skype with our friends from other countries, we share photos, we Tweet, we check-in our favourite places. We sometimes depend too much on technologies but then feel violated and powerless when we hear talking about Big Data and how everything that we do is monitored.

"The digital traces we leave behind each day reveal more about us than we know. This could become a privacy nightmare-or it could be the foundation of a healthier, more prosperous world" (Alex Pentland, 2013)

Chapter 2

2.1 The Origins

As stated in the introduction, it is not simple to define Big Data as it is something so broad and diverse. To simplify, we could define it as a collection of dataset (all the sets of information composed of separate elements) which gather very heterogeneous information able to communicate between each other and that give access to an immense amount of information. These information are then used to compute complex analysis overlaying and confronting as much data possible to obtain a clear vision of the studied phenomenon (for example a financial study). So, the information does not come from a database or a dataset , but from a larger entity which contains them both and which makes different types of data accessible: emails, GPS locations, images, archives, information gathered from social networks.

Data has been collected and analysed for many centuries. Man has always been doing this since he invented writing in ancient Mesopotamia so to have the possibility to keep record of all information. If we consider Egyptians, Chinese, Romans, Greeks, they all conducted censuses. We can also recall the "Doomsday book" which was one of the first historical public record compiled in 1085-1086 containing all legal information of that period (English people, land and properties). Of course these methods were slow and costly and it was only in 1662, that John Graunt, an English draper, came up with a new method to study mortality in London due to the great Plague. He is considered being the founding father of modern statistics. In an article in the dictionary of Scientific Biography, Frank N. Egerton wrote: "(Graunt) introduced the use of statistical samples [though he] did not pursue this subject far enough to determine the sizes of samples or means of selection needed for insuring accuracy. (Graunt) also realized that demographic procedures could be used to make projections concerning both past and future populations." So, even though this innovation was a real revolution, data collection was still far from being accurate and still required a lot of time and money; in

fact censuses were not executed often. In 1881, a new and much more efficient technique was introduced by Herman Hollerith. He designed a punch-card tabulation machine to collect census data for the U.S. Census Bureau. This new machine managed to conclude in only one year what before would have taken almost 10 years of hand tabulation. In 1896 Hollerith founded the Tabulating machine Company and sold his invention which became part of IBM in 1924.

Another interesting example of innovative and useful data collection is the one of an American librarian, J. Edgar Hoover who became the first Director of the Federal Bureau of Investigation (FBI) in 1924. He introduced new methods (some illegal) to collect information and evidence such as a centralized fingerprint file and forensic laboratory.

Later on, in 1934, random sampling was introduced by Jerzy Neyman, a Polish statistician. This has been the real innovation in the science of statistics but, if we want to dig deeper and have more accurate information, random sampling is still not good enough. Most of the times though, sampling is the only affordable method yet. In the following chapter we will understand how Companies can use Big Data effectively such as Google, who analysing search queries, understand and detect flu trends or how retailers increase their sales by positioning products in their stores.

2.2 The Story of a New Era

Big Data can be identified as the beginning of a new transformation; MIT's Technology Review stated that it is one of the "10 Emerging Technologies that will change the world" together with wireless sensor networks, injectable tissue engineering and nanoimprint lithography. The innovation is in the optimized use of statistical methods and in the rapid storage growth and, as any other new and powerful technology, it leads us to boundless possibilities if used in the right way. Of course, if used by wrong people, with bad intentions, it might cause unimaginable damage (which we will discover in chapter 3). In fact, before Big Data can be used efficiently, it must be filtered and managed through data analytics. This process requires high-performance technologies such as in-database analytics which move important data to where it has to be located or in-memory analytics, able to solve and provide answers to highly complex

problems rapidly. Subsequently, the analysed data is ready for organizations to be used for specific purposes also using recent technologies such as ample storage, fast processors and cloud computing.

"Every two days now we create as much information as we did from the dawn of civilization up until 2003. That's something like five Exabyte of data.

Let me repeat that: we create as much information in two days now as we did from the dawn of man through 2003."[1]

Here are some examples of how fast data grows: Google processes more than 24 petabytes of data each day, Facebook has more than 10 million photos uploaded every hour and YouTube gets an hour of video uploaded every second. To give a visual idea of how much data is around us, we can imagine 52 layers of books coving the surface of the United States.

Today, everything moves faster, everything is easier and most of the time we take these for granted. Think of how using Skype on your laptop for a conference call changed our life: we can work from our home, waste less time in traffic and spend it with our family. If we look back, not 50, but only 10 years ago, we notice the huge difference in lifestyles. This can definitely be considered as being one of the fastest revolutions ever happened.

[1] Eric Schmidt, 2010

Ride the Big Data Wave

90%

of the world's existing data has been created in the last

2 YEARS

By 2020

35 Zettabytes worth of data will be created

*35 Zettabytes = 35,000,000,000,000,000,000,000 bytes

1/3 of that data will be stored in or will have passed through the cloud

10,000

Libraries of Congress worth of data are created annually by U.S. companies

80,000 km

is the distance past the moon the current global storage capacity for digital information stored on a CD stack would stretch

Every 2 days as much information is created as from the dawn of civilization up until 2003

There are actually more pieces of digital information than stars in the universe

EVERY DAY

1 Billion	2.5 Quintillion	8 Years
		YouTube
pieces of content are shared via Facebook's Open Graph	bytes of data are generated by people	of video is uploaded to YouTube

De-risk Your Database Vendor Selections by Aligning with Market Trends

See the Market Trends with the Progress 2014 Data Connectivity Outlook

An Uninformed Database Vendor Selection can be Costly to Your Business

*Source: http://goo.gl/omNJuY

Full Paper: http://goo.gl/yxp9DI

Chapter 3

How Big Data Affects the World

3.2 Big Data Everywhere

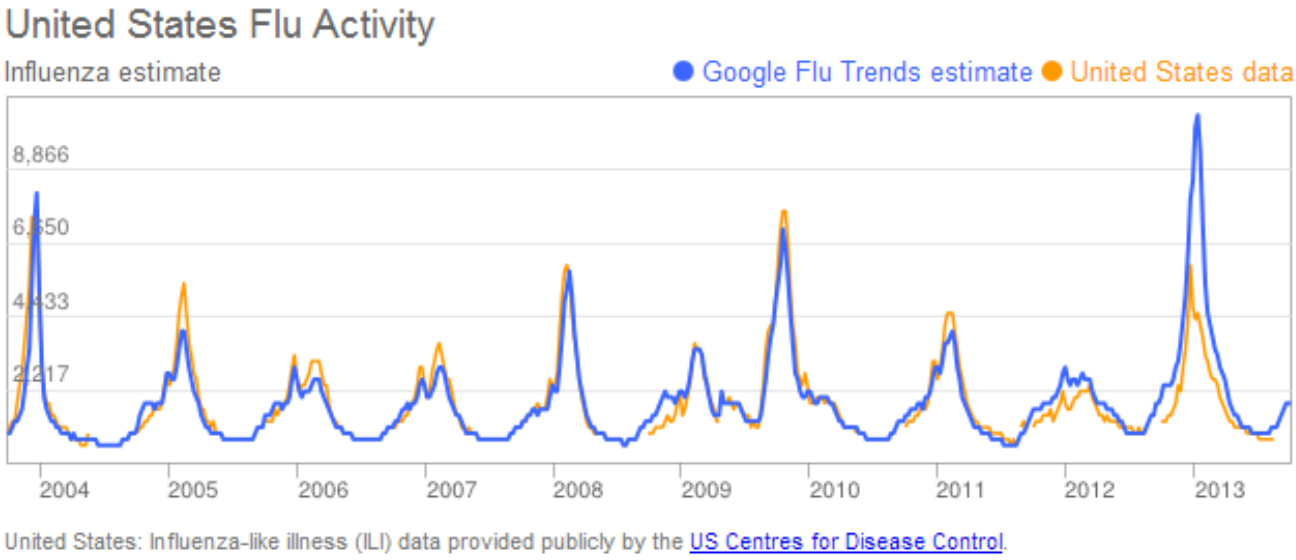
Nowadays almost every sector, retail, finance, healthcare, crime, education, is using Big Data as their key to success. There are unimaginable ways in which technologies can help us: from small daily things such as reminding us of our meetings, to much more important situations like preventing a car accident or monitoring a person's pacemaker. We will now discover how Big data is used to improve healthcare, prevent crime and increase marketing effectiveness.

3.2 Big Data and Healthcare

Big Data plays a fundamental role in the health sector. An astonishing example is the one of Google being able to formulate predictions on flu by analysing what people searched for on their computers. *"About 90 million American adults are believed to search online for information about specific diseases or medical problems each year, making web search queries a uniquely valuable source of information about health trends."* [2]

[2] Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. "Detecting Influenza Epidemics Using Search Engine Query Data." Nature 457.7232 (2008): 1012-014. Web.

Accurate studies have evidenced that there is a close relationship between people searching for flu information and the number of people actually having the symptoms. Using elaborate algorithms, Google was able to track where the flu was spread and consequently could use the information also to predict when it would spread in the near future. In the 2009 H1N1 pandemic, Google Flu Trends managed to retrieve useful information before the CDC (Centres for Disease Control and prevention) released its



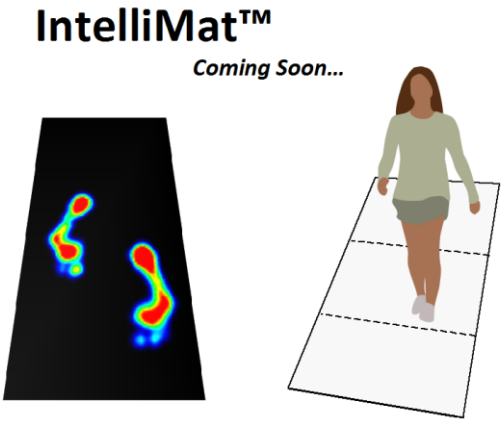
report. Although Google's information has been extremely useful, there have been many concerns about privacy matters. People worry that through their search queries, their privacy has been violated but what is important to point out is that the database for this project retained no private information about the identity of the person using the computer.

The United States Department of Defence's Electronic Surveillance System for Early Notification of Community-based Epidemics (ESSENCE), monitors all available health data to predict epidemic and biological outbreaks. Born in 1999, it now has about 400 facilities all around the world monitoring the Military Health System. What Essence does is to detect the first signs of illnesses from clinic reports, doctors and pharmacies and breaks down all the information to find similar patterns. Once the similarities have been spotted, Essence can predict where and when the illnesses could occur.

"We detected a gastrointestinal outbreak in Korea," Mansfield says. "I called my boss, and he asked me, 'When did it happen?' "Korea is 13 hours ahead of Washington. So Mansfield simply answered: "Tomorrow."[3]

Other examples of how Big Data is used in healthcare are GPS-enabled trackers that monitor the use of inhalers by asthmatics or Wi-Fi pacemakers which allow doctors to monitor the patients anytime and anywhere. These are only two out of all the millions of already existing or possible uses of Data-driven technologies in this sector.

One of the most recent data-based technologies is a pressure-sensitive floor mat which is able to give health details about the person walking on it. The IntelliMat is the newest of the Tactonic Technologies' inventions and it completely relies on cloud-based analytics and data analysis. Depending on the pressure of footsteps, the square IntelliMat tiles memorizes weight, step interval and length. It then saves the data in the cloud system and it compares the new data to analyse the person's activity and health and give them feedback. Tactonic claims that the IntelliMat is able to determine whether a person might have arthritis or Parkinson's disease so having them in hospital could be very useful too. These tiles could also be useful in stores to track the shoppers' behaviour through stores and provide an important marketing tool.



3.2.1 personalised drugs

The concept is straight-forward: when an oncologist gets a tumour biopsy, he analyses the genome sequence and find the mutations in the cancer cells which enables the doctors to create personalised treatments. All the data collected makes it easier to simulate the patient's reaction to certain drugs; not everyone's response to a treatment is the same. This flow of information has resulted in being a very effective tool in healthcare: doctors can make better diagnoses and give better preventions to their

[3] Sharon Weinberger, 2008

patients by monitoring them with smartphone trackers, real time monitoring and sensors. The pharmaceutical industries in this way, also reduce their drug failure costs because they create specific treatments which have a very low probability of being ineffective. Of course, there are other type of costs such as the analysis of each patient's dna but they are still much less than more general costs needed to create a common cancer drug. In the first place, patients have reduced side effects and have improved drastically their quality of life. Also, the healthcare system has been saving about $300 billion dollars every year and thousands of life. This can indeed be regarded as one of the most effective ways of using Big Data.

3.3 A Minority Report Future?

"This is the next era of policing… very soon we will be using a predictive policing model where, by studying real-time crime patterns, we can anticipate where a crime is likely to occur."[4]

Big Data can stop crime before it even happens. This may seem like a science fiction film plot as the 2002 film Minority Report directed by Stephen Spielberg, in which data scientists are able to predict when and who will commit a crime.

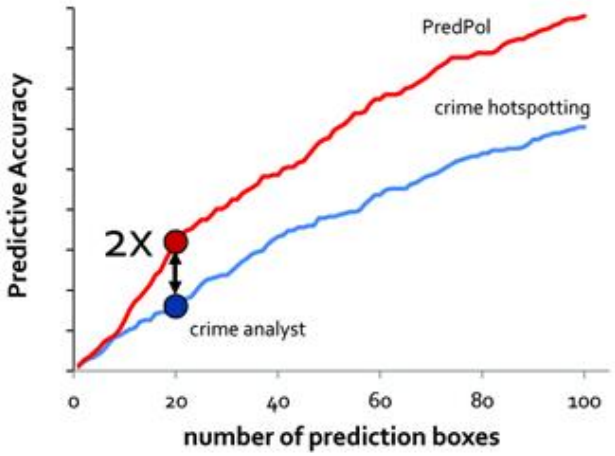
"In the year 2054, crime is virtually eliminated from Washington D.C. thanks to the Pre Crime elite law enforcing squad. They use three gifted humans, the precogs, with special powers to see into the future and predict crimes beforehand."[5]

Also, in the TV series "Person of Interest", a "machine" predicts who is going to die within 28 hours leaving it to a data scientist and a professional soldier to save them.

Only about ten years later (and not fifty two) a few police departments started using Big Data to solve many crimes. In Memphis, the Blue CRUSH (crime Reduction Utilizing Statistical History) program helped crime drop by about 30%.

[4] William Bratton, Active Commissioner of NYPD and former Chief of LAPD
[5] Yves de Montcheuil, 2012

The Chicago Police Department is also using its own crime predicting program. In their database, they have locations where crimes frequently occur, and phoned incidents to map the criminal "hotspots".



The Predpol (predictive policing) is a crime predicting program used by the Los Angeles Police Department and its mission is*: "place officers at the right time and location to give them the best chance of preventing crime."* Its computations are based on the same predicting models for earthquake aftershocks and is able to anticipate where and when a crime is most likely to happen giving very accurate "prediction boxes" of about 500 x 500 feet. Using Predpol, predictions have been twice as accurate as the ones using the traditional hotspot mapping system. Within only six months of introducing Predpol

in the area of Foothill in Los Angeles, there has been a 12% fall in property crimes compared to the previous year (2010). Up until now, these predictions have been mostly used for thefts and burglary because these are the crimes which present most past data of all but soon, with more information, also other types of crimes can be predicted too.

In these movies, as it happens in real life, data is collected from all sorts of sources: cell phones, cameras, GPS, street webcams and text sources which are all stored and analysed for different purposes. It can be quite disturbing but we are constantly monitored by all these devices and most of the times we are the ones asking or accepting it. Each time we download an application on our cell phone, we accept the terms and conditions of that app which most of the time asks us to see our social network information or GPS settings.

3.4 Big Marketing

It is almost Christmas, one of the busiest shopping periods of the year in New York and a type of cell phone is being sold like hot cakes. A few years ago, the shop who reacted to the demand in the fastest way obtained the advantage but today Big Data is the key to gain competitive advantage. Retailers can now predict future trends and respond promptly the costumers' demands. So, if a shop wants to sell more cell phones than its competitors', it will start working in advance to predict the consumers' needs and when they are most likely to purchase the good. What Big Data does is to analyse data from web browsing, social media or cell phone industry data and create predictive trends. Once the statistics have been done, the shop can order the right number of cell phones, give them an "attractive" price and be ready to satisfy their consumer's needs. Lastly,

once the shop has gained customer loyalty, they can capture them with more personalized offers and keep in touch with them through text messages, emails or mail.

When we go to a shop and use our credit card and/or fidelity card, we automatically give detailed information about our buying habits to the retailer. These details are essential for the company to have a more personalized connection with each consumer. In this way, the retailer knows what we like most, where and when we prefer buying specific products.

Unfortunately there have also been some negative outcomes for the excessive use of data collected. The American discount retailer Target is able to understand when a woman is pregnant simply by analysing their consumption patterns but this could sometimes invade people's privacy like in the case of a man discovering that her daughter was pregnant because Target was sending her coupons for baby products at home, and of course he had much to complain about this. Andrew Pole was the first Target statistician who managed to find a relation between what women bought and their pregnancy.

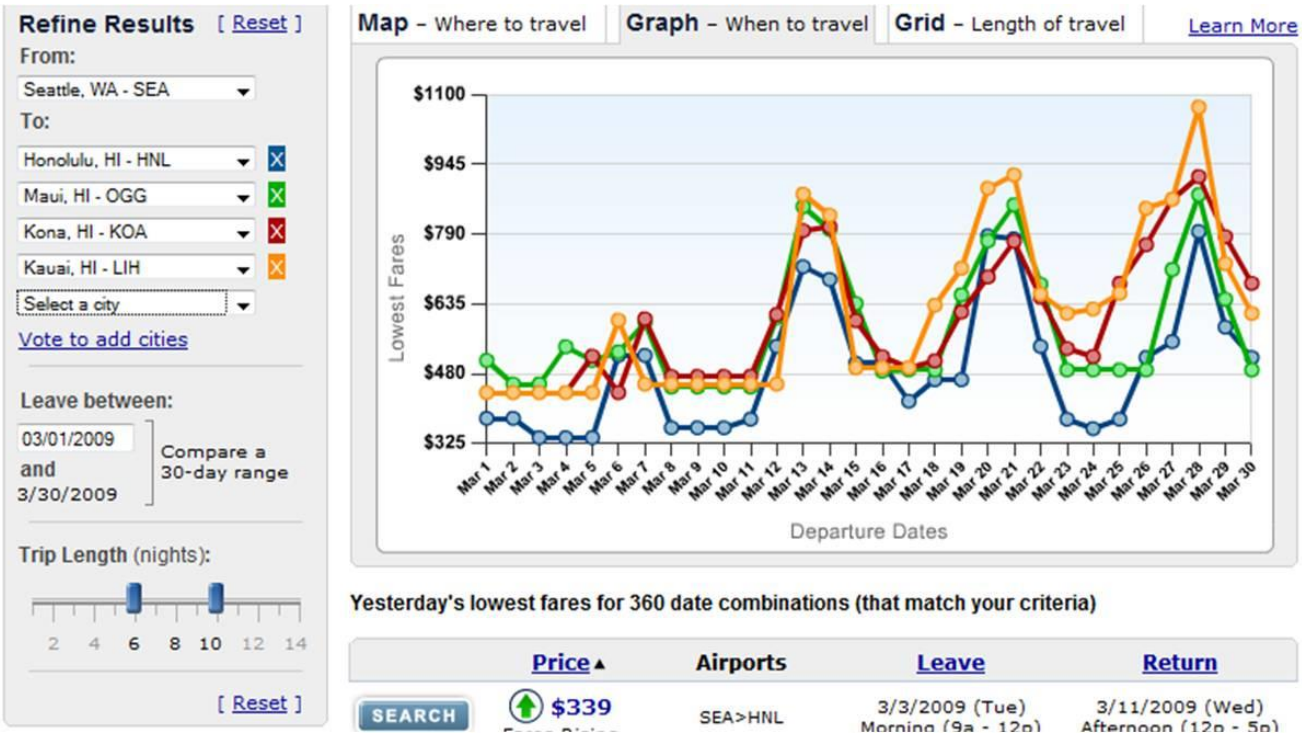
"[Pole] ran test after test, analysing the data, and before long some useful patterns emerged. Lotions, for example. Lots of people buy lotion, but one of Pole's colleagues noticed that women on the baby registry were buying larger quantities of unscented lotion around the beginning of their second trimester. Another analyst noted that sometime in the first 20 weeks, pregnant women loaded up on supplements like calcium, magnesium and zinc. Many shoppers purchase soap and cotton balls, but when someone suddenly starts buying lots of scent-free soap and extra-big bags of cotton balls, in addition to hand sanitizers and washcloths, it signals they could be getting close to their delivery date."[6]

In 2007, Howard Schultz, Starbuck's CEO, applied a data-driven approach to improve the company's growth plan. The use of Esri, a mapping software, helped determine the best places to open the next Starbuck's store. "The software is also helping to determine where the next 1,500-plus stores should be placed not only to help the company expand,

[6] Charles Duhigg, the New York Times, 2016

but drive revenue for new store developments."[7] Also, with all the data collected from its customers' loyalty cards, Starbucks can identify each customer and know what their favourite products are, where and when they usually buy them: a full buying behaviour activity. This allows Starbucks to have detailed customer's profiles and target their advertising. Lately though, thee personalised advertisements are not aiming at offering discounts to loyal customers but, on the other hand, to customers whose buying behaviour shows that they might not come back so to keep them wanting to return.

Another business sector which has been adopting the use of Big Data is the airline industry one, in particular selling flight tickets. Oren Etzioni, an American computer scientist, in 2003, while travelling by plane, noticed that his seat neighbours all paid different fares for the same flight. He found this completely unfair and immediately got to work to create a flight fare predictor, Farecast. The program was able to anticipate if fares would increase or decrease by analysing the past changes in prices. Consumers



were now able to go on the website and choose whether to buy their tickets now or later. By 2008 Etzioni also wanted to add predictions for hotel rooms, car rentals and concert tickets but before he managed to do so, Microsoft bought Farecast for $115 million making it become part of his search engine, Bing. Now, Bing doesn't offer price

[7] Howard Schultz, 2008

predictions anymore but only simple travel information. Was it maybe too consumer friendly? On the other hand, Google has started using flight fares predictions too with "Google flight" and, also regarding Etzioni, it is much more innovative than Bing; in fact he has been very disappointed by how Bing has not been able to take advantage of his own program.

Chapter 4

Privacy issues

4.1 Balancing the pros and cons

As for every revolution, it takes some time to accept and understand it completely. The rapid growth of technology has made it possible to collect and analyse enormous amounts of data. In the previous chapters we found many possible positive outcomes of using Big Data in various sectors.

One of the biggest privacy issues is that data should not be used for secondary purposes without the " owner" not knowing. This can be found in the 2012 FTC's Privacy Report which states:

Safeguards for personal privacy based on our concept of mutuality in record keeping would require adherence by record-keeping organizations to certain fundamental principles of fair information practice.

• There must be no personal-data record-keeping systems whose very existence is secret.

• There must be a way for an individual to find out what information about him is in a record and how it is used.

• There must be a way for an individual to prevent information about him obtained for one purpose from being used or made available for other purposes without his consent.

• There must be a way for an individual to correct or amend a record of identifiable information about himself.

• Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take reasonable precautions to prevent misuse of the data.[8]

[8] FTC , 2012 Privacy Report

The problem is that Big Data implies analysing "hidden" data often obtained from companies who then sell all their collected information to other companies. A large number of people are very concerned about the private information collected and used for unclear purposes. In these cases, people feel they have no control of how their personal details are used and feel violated. We are constantly monitored, we seem to be living in "The Truman Show" in which our every move is captured by cameras and sensors and the idea of being punished for a crime that has still not happened can seem quite scary.

It is very interesting to see the different approaches and attitudes towards these privacy issues between Europe and the US. In Europe people are much more diffident and do not seem to think that Big Data can bring them any kind of benefits. On the other hand, Americans are optimistic and believe that data collection can help improve many aspects of their lives. At the same time though, Americans are much more conscious about all the risks and are therefore able to protect themselves than Europeans. What should be done is to keep people more informed on how and why companies collect data especially in cases as the Google flu trends, crime prediction or marketing campaigns. Another way to solve the problem would be to give people the ownership of their data. Alex Pentland used the Old English Common Law for the basics of ownership: right of possession, use and disposal.

1. You have a right to possess your data. Companies should adopt the role of a Swiss bank account for your data. You open an account (anonymously, if possible), and you can remove your data whenever you'd like.

2. You, the data owner, must have full control over the use of your data. If you're not happy with the way a company uses your data, you can remove it. All of it. Everything must be opt-in, and not only clearly explained in plain language, but with regular reminders that you have the option to opt out.
3. You have a right to dispose or distribute your data. If you want to destroy it or remove it and redeploy it elsewhere, it is your call.[9]

Big companies like Google and Facebook are so powerful because they own massive quantities of personal information and by owning them, they get to control everything. In this "information" era, data collected from social networks and other devices has become the new money: information is synonym of power. All the data collected is hidden from us and the only ones who have control over it are the companies owning it. It is not hard to imagine how things could go wrong, if the information is used by the wrong computer scientists, we could end up in a science fiction situation in which we are controlled by unknown entities. Unfortunately, there is little science fiction in this because in some laboratories such as the one in the University of California, Bakerly, scientists have already managed to enter in people's minds, read them and influencing their decisions: all this just using Big Data statistics. Information is free, but if we started to own it, as we said before, and sold it, we cannot say whether this would really help solve the problem because people would start giving different prices to their own information and take advantage of this ending up in a black market of information.

[9] Alex Pentland, the Global Information Technology Report 2008-2009

Conclusions

It is hard to say if the use of Big Data will lead to a better future; wheat we can say is that up until now, it has had more of a positive impact than a negative one. By creating multiple connections with all the Data, scientists are able to study and prevent many diseases, the police can prevent thefts and crimes, shops can increase their sales and many other things. The fundamental thing is that whatever data is collected, it must be used in the right way and by the right people for important purposes without damaging the owners. If, in the future we are not able to control all the data flow, there might be disastrous outcomes.

Bibliography

"10 Emerging Technologies That Will Change the World - MIT Technology Review." MIT Technology Review. Web. 17 June 2014.

Arthur, W. Brian. The Nature of Technology: What It Is and How It Evolves. New York: Free, 2009. Print.

"Assessment of ESSENCE Performance for Influenza." Centers for Disease Control and Prevention. Centers for Disease Control and Prevention, 08 Apr. 2011. Web. 23 June 2014.

"BIG DATA." A Day in Big Data. web. 22 June 2014.

Big Data Now Current Perspectives from O'Reilly Media: 2012 Edition. Sebastopol, CA: O'Reilly Media, 2012. Print.

"The Data-Driven Society." Web.

EMCVoice. "How Big Data Is Transforming Drug Development." Forbes. Forbes Magazine, 17 Mar. 2014. Web. 26 June 2014.

"The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." 23 June 2008:Web. <http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory>.

Ghemawat, Sanjay, and Jeffrey Dean. MapReduce: Simplified Data Processing on Large Clusters (n.d.): n. pag. Google Inc. Web.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. "Detecting Influenza Epidemics Using Search Engine Query Data." Nature 457.7232 (2008): 1012-014. Web.

Hays, Constance L. "What They Know About You." The New York Times. The New York Times, 13 Nov. 2004. Web. 13 June 2014.

"Health.mil." Health.mil - Web. 23 June 2014.

"How Big Data Can Transform Society for the Better." Scientific American Global RSSWeb. 21 June 2014.

"The Human Face of Big Data." The Human Face of Big Data. Web. 20 June 2014.

I Number 1 I Spring 2, Richard Schmalensee. "Jeffrey Rohlfs' 1974 Model of Facebook: An Introduction Web.

"J. Edgar Hoover: The Librarian Who Became America's 'most-powerful Man'" Constitution Daily. Web. 17 June 2014.

Kearney, Tony. Who Owns the Future? Milton Keynes: AuthorHouse, 2008. Print.

Lanier, Jaron. Who Owns the Future?. Print.

Lanier, Jaron. You Are Not a Gadget: A Manifesto. New York: Alfred A. Knopf, 2010. Print.

Mayer-Schönberger, Viktor, and Kenneth Cukier. Big Data: A Revolution That Will Transform How We Live, Work, and Think. Boston: Houghton Mifflin Harcourt, 2013. Print.

Needham, Jeffrey. Disruptive Possibilities: How Big Data Changes Everything.. Print.

Ohlhorst, Frank. Big Data Analytics: Turning Big Data into Big Money. Hoboken, NJ: Wiley, 2013. Print.

Pentland, Alex "Sandy". "The Data-Driven Society." Scientific American 309.4 (2013): 78-83. Web.

"Personalized Medicine: Taming the Big Data Ocean." Phenomena Personalized Medicine Taming the Big Data Ocean Comments. Web. 26 June 2014.

PlanetVoice, IBM Smarter. "Five Ways Companies Can Compete Using Big Data and Analytics." Forbes. Forbes Magazine, 14 Apr. 2014. Web. 20 June 2014.

Press, Gil. "A Very Short History Of Big Data." Forbes. Forbes Magazine, 09 May 2013. Web. 17 June 2014.

"Ride the Big Data Wave." Data Connections. Web. 21 June 2014.

Santovena, Alejandro Zarate. Big Data: Evolution, Components, Challenges and Opportunities -2013. Print.

Smolan, Rick, and Jennifer Erwitt. The Human Face of Big Data. Sausalito, CA: Against All Odds Productions, 2012. Print.

Snickars, Pelle, and Patrick Vonderau. Moving Data: The Iphone and the Future of Media. New York: Columbia UP, 2012. Print.

"Top Five Reasons Big Data Is Advancing Personalized Medicine." Top Five Reasons Big Data Is Advancing Personalized Medicine. Web. 26 June 2014.

Zimmer, Carl. "Enlisting a Computer to Battle Cancers, One by One." The New York Times. The New York Times, 26 Mar. 2014. Web. 26 June 2014.