DPP Open Thoughts Paper N°19

11 March 2014

Big Data: A Blessing or a threat to Human Rights?

1. Introduction

The phenomenon of 'Big Data' is showing an ever growing impact on sociology and political science. It has a tangible influence on our daily lives and the functioning of not only commercial but also political institutions. We are accustomed to the fact that Amazon can e-mail its customers lists of selected books that they might find of interest, that LinkedIn publishes certain job offers on one's personal LinkedIn page for one's attention, and TomTom manages to alert one to roadblocks, diversions and traffic congestion? Is it possible to conceive, in the near future, that Twitter might be able to predict the outcome of national elections, or that Google could provide a warning about the next H1N1 pandemic[1]?

In their daily lives, people are increasingly assailed by information. '***Datafication***'[2] has become a widespread and high potential issue. It is caused by a number of factors, which includes the public embracing the social media, the digitalisation of books, music and DVD / films, the intensification of use of the Internet as well as cheaper and better sensors that allow us to measure and track everything. It is also the result of human activity as a whole that leave small digital traces without any awareness on the part of the individual. This gigantic source of information has reached colossal volumes. The new data produced over the last two years is the equivalent to the amount of data produced in the past 2000 years and it continues to grow exponentially. This 'data-mountain' is perceived as being a goldmine for businesses, because it seems to be possible to retrieve specific data in real time by means of new techniques[3].

2. The 'Big Data' Phenomenon

Big Data[4] is a 'buzzword' used to describe a **massive volume of both structured and unstructured data** that is so vast and complex that it becomes difficult to process through the use traditional database management tools and data processing applications. In most business scenarios, the data is too big, or it moves too fast, or it exceeds processing capacity. The exponential growth in size is caused by information-sensing tools (mobile devices, aerial sensory technologies (via remote detection), software, cameras, microphones and wireless sensor networks).

---

[1] Viktor Mayer-Schönberger & Kenneth Cukier; '*Big Data: A Revolution That Will Transform How We Live,Work and Think*'; www.hmhbooks.com, 2013, 242p.

[2] '**Datafication**' means the conversion of human activity (especially social activity) into data format. See Kenneth Cukier; '*The Birth of Datafication'*, video (ref. http://bigthink.com/videos/the-birth-of-datafication).

[3] For example, Hadoop or NoSQL.

[4] '**Big Data**' is a term created in 2005 in order to enclose the trend of information and data accumulation.

While the term may seem to refer to the volume of data[5], it is increasingly used in a broader sense. The term Big Data[6], especially when used by traders, may refer to the **technology** (which includes tools and processes) that an organisation needs to handle the large amounts of data and storage facilities, requiring massively parallel software and servers.

The aim of '***Big Data management***' is to ensure a high level of data quality and accessibility for appropriate strategic analysis. Businesses, government agencies and other organisations employ Big Data management strategies to help them to cope with fast-growing pools of data and draw conclusions for their corporate strategies. Effective Big Data management helps companies to locate valuable information in large sets of unstructured data and semi-structured data from a variety of sources, including call detail records, system logs and social media sites, such as Facebook and Twitter.

'Big Data analytics' is the process of examining large amounts of data of a variety of types to uncover hidden patterns, unknown correlations and other useful information. Such information can provide competitive advantages over rival organisations and result in business benefits, such as more effective marketing and increased revenue. In politics, it may help to better understand voter's preferences and neutrality.

3. Can Big Data Change Society?

In a recent debate on the social impact of Big Data[7], Viktor Mayer-Schönberger and Kenneth Cukier emphasise the potential of Big Data analysis by means of '***data mining***'[8] in order to shape every part of society from health care and education to urban planning, the protection of the environment or climate change. Firstly, capturing much more data on certain (social) phenomena provides more necessary detail than was the case previously when a sample of data did not lead to effective results. Secondly, some 'messiness' in the data analysed is accepted because of the amount of data available without worrying about the exactness of the results. Finally, highly valuable and previously unknown connections and correlations between different information sources lead to predict future tendencies (and problems) more carefully and accurately.

---

[5] Although 'Big Data' does not refer to any specific quantity, the term is often used when discussing 'petabytes' (1 million of Gigabytes) and even 'exabytes' (1 billion of Gigabytes) of data (year 2012).

[6] The criteria that define Big Data (Volume, Variety, Velocity and Veracity) are the current four '4V's' or 'properties' or 'dimensions' of Big Data. *Volume* refers to the amount of data, *variety* refers to the number of types of data while *velocity* refers to the speed of data processing. Finally, *veracity* refers to the messiness of the data.

[7] Viktor Mayer-Schönberger & Kenneth Culier, '*Big Data's Bright and Dark Sides*', in Debates & Series, Skoll World Forum debates: 'How Can Big Data Have a Social Impact?' (ref.: http://skollworldforum.org/2013/03/19/big-datas-bright-and-dark-sides/), 19 March 2013.

[8] '**Data mining**' (sometimes called 'data-' or 'knowledge discovery') is the process of analysing data from different perspectives and summarising it into useful information - information that can be used to increase revenue, cut costs, or both. Data mining software is one of a number of analytical tools for analysing data; it allows users to analyse data from many different dimensions or angles, categorise it, and summarise the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. For a more extensive explanation, see Toon Calders & Bart Custers; 'What Is Data Mining and How Does It Work?', 2013, 42p. (ref.: http://link.springer.com/chapter/10.1007%2F978-3-642-30487-3_2#).

In addition to this, the potential of Big Date affects **governance**; building systems and using technology that makes information by and about governments more accessible and exploitable, can improve government governance and amplify the power of engaged citizens, to the benefit of society. On the contrary, the situation could become prejudicial should government agencies or even private (profit) organisations use Big Data for less noble purposes, such as for surveillance actions, or simply for the benefit of large profits. From that moment on, the protection of privacy and human rights is at risk[9].

4. The Dark Side of Big Data: When the Protection of Human Rights is at Stake

Much of what constitutes Big Data is information about people. Collecting data by means of so-called '***predictive analyses***' on personal data is a typical form of mass-manipulation of information. Routine Big-Data analysis can effectively manicure personal data by creatively collating scattered pieces of data about changes in human behaviour in order to predict future needs[10]. While Big Data may produce benefits for development initiatives, it also presents serious **risks**, which are often ignored. In pursuit of the promised social benefits that Big Data may bring, it is crucial that fundamental human rights and ethical values are not set aside.

Once data has been collected, there is absolutely no control over who uses it or how it is used. If private sector data falls into the 'wrong hands', it could enable the **monitoring of individuals**, their identification and their surveillance. Despite guarantees of anonymisation, the correlation of separate pieces of data can (re)identify an individual and provide information about them that is even more private than the data they consented to share, such as their religion, ethnicity or sexual orientation. If this were to occur in certain contexts the consequences could have a tragic impact, especially should the data concerned relate to vulnerable groups such as minorities or refugees, as well as societal groups including journalists, social dissidents and human rights advocates. The threat becomes even more serious when data collection is conducted in the name of safety, especially by governments. By means of '***profiling***'[11], it is possible to detect so-called 'outliers' within a group and act against individuals who demonstrate behaviour that is outside the expected norms within a community, but which could jeopardise fundamental rights due to the risks of profiling individuals with regard to discrimination, inequality, stereotyping, stigmatisation and inaccuracy of the decision-making process [12].

Because Big Data is derived from aggregated data from various sources (which are not always identifiable), there is no mechanism to request the **prior consent** of a person for the resulting data that emerges. In many cases, this kind of data is more personal than the set of data to which the person has given their consent. Currently, data users must give individuals notice and ask them to

[9] For exemple, the US 'PRISM' data mining programme, reveiled by Edward Snowdon.

[10] A well-known example of '**predictive analysis**' is the change in shopping habits that predict the delivery date of pregnant shoppers so that they can be targeted for baby-related advertisements.

[11] '**Profiling**' can be described as 'an automatic data processing technique that consists of applying a "profile" to an individual, particularly in order to take decisions concerning her or him or by analysing or predicting her or his personal preferences, behaviour and attitudes'. See CM/Rec(2010)13, 23 November 2010.

[12] As there are: the right to privacy and protection of personal data; the principle of non-discrimination.

consent at the time of the collection of personal information. But this mechanism fails in the Big Data era when the value of data is in its secondary uses, and thus without the individual's consent. In this context, personal data analysis could lead to **damaging an individual's integrity**, or having serious consequences on a person's private or professional life[13]. Today, it seems almost impossible to redress the harm caused by this kind of data gathering because it is operating outside of current privacy protection mechanisms.

A final dilemma is one that is not unique to Big Data, but one that society needs to vigilantly safeguard against, which is namely the '**dictatorship of data'**[14]. Affording Big Data a special status and endowing it with more meaning and importance than it justly deserves, results in foregoing judgment and common sense. As Big Data starts to play a part in all areas of human behaviour, this trend to place undue trust in the data may only grow.

5. Big Data: a Threat to Human Rights or to the Benefit of Mankind?

Some specialists believe that Big Data and the concern for privacy can co-exist. In the aftermath of the NSA scandal over the collecting of random information, Dr. Alexander Dix, Berlin Commissioner for Data Protection and Freedom of Information, believes that there are possible ways of **regulating for privacy protection** in the future. Firstly, there should be an international agreement on where to set the limits on monitoring internet traffic and people's behaviour, on what kind of data processing should not be allowed under any circumstances. On a regulatory level, one should work for international guarantees of privacy, such as in the UN Covenant on Civil and Political Rights. Secondly, technical solutions like improved tools for self-protection should empower individuals to protect their own communication.

'***Pseudonymisation***', Dr. Dix adds, is a method to reduce the personalisation of data. This important tool of systemic data protection is recognised in German law and the European Commission has agreed to take this proposal on board.

6. Government Initiatives: too limited and too late?

Major events with effects on a global scale require efficient and effective responses. So did the Snowdon-revelations about the PRISM-programme which has had many repercussions. The **White House** has now launched an action plan on the collection, availability and the growing use Big Data analytics and its potential **impact on the future of privacy**. As part of the program, the US President's Council of Advisors on Science and Technology (PCAST) will examine an in-depth study to explore the technological dimensions of the juncture of Big Data and privacy[15].

---

[13] For example, analysis of a potential health issue of an individual may have an impact on the employment or health insurance prospects (with a correct inference) or might lead to discrimination in the workplace (with an incorrect inference).

[14] Viktor Mayer-Schönberger & Kenneth Cukier; 'Big Data's Bright and Dark Sides'; Skoll World Forum debates: 'How can Big Data Have a Social Impact?', l.c.

[15] Ref. : http://cloudtimes.org/2014/01/29/us-government-starts-big-data-review-to-get-privacy-right/

At **European level**, at the beginning of 2012, the European Commission launched a major reform of the EU legal framework on the protection of personal data. The '**General Data Protection Regulation**' (GDPR) will strengthen online privacy rights and protect citizens against the use of personal data that could be processed for illegal purposes, and provides rules for harmonising all data protection regulations throughout the European Union. This makes it thereby easier for European and non-European companies to comply with the new regulation[16]. Besides the new Regulation, a Directive will protect personal data processed for the purpose of prevention, detection, investigation or prosecution of criminal offences and related judicial activities[17].

In recent years, the **Council of Europe** undertook new efforts to protect against data-based activities that affect essential and **fundamental human rights.** As new techniques to explore Big Data appeared, several measures, resolutions and recommendations were discussed in order to provide member states with a necessary legal basis upon which they could focus domestic legislation. The Committee of Ministers of the Council of Europe, therefore, in recent years adopted several recommendations in order to preserve these fundamental rights, by focusing especially on legally safeguarding the access to internet users' personal data by search engines, informing search engine users about possible threats related to data use, and urging member states to enforce compliance with the applicable data protection principles[18].

The abuse of personal data in the context of the above-mentioned 'profiling' runs counter to fundamental values (democracy and the rule of law, autonomy and self-determination) and rights (the right to privacy, data protection and non-discrimination)[19]. On 23 November 2010, the Committee of Ministers of the Council of Europe adopted a **Recommendation** to member States 'on the protection of individuals with regard to **automatic processing of personal data in the context of profiling**', and determined the general principles with regard to the respect of the afore-mentioned values and rights, the conditions under which personal data processing in the context of profiling should take place and the rights of the data owners[20].

Last but not least, on 8 November 2013, the Ministers of the Council of Europe States declared their strongest support for the **right of freedom of expression and the right to private life** against all sorts of threats against in the current digital age, including data collection for surveillance purposes undermining or even destroying democracy[21].

---

[16] The GDPR will replace the current Directive 95/46/EC in the course of 2015, after a 2-year trial.

[17] See also http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm.

[18] Recommendation CM/Rec(2012)3 of the Committee of Ministers to member States 'on the protection of human rights with regard to search engines', 4 April 2012 (ref.: https://wcd.coe.int/ViewDoc.jsp?id=1929429).

[19] See Valeria Ferraris, and Francesca Bosco and Elena D'Angelo, o.c., p.13-17.

[20] Recommendation CM/Rec(2010)13 of the Committee of Ministers to member States 'on the protection of individuals with regard to automatic processing of personal data in the context of profiling', 23 November 2010 (ref.: https://wcd.coe.int/ViewDoc.jsp?id=1710949).

[21] Council of Europe Conference of Ministers responsible for Media and Information Society; 'Belgrade Declaration on Freedom of Expression and Democracy in the Digital Age', Belgrade, 8 November 2013 (ref.: http://www.coe.int/t/dghl/standardsetting/media/Belgrade2013/Belgrade%20Ministerial%20Conference%20Texts%20Adopted_en.pdf).

7. Observations

The growing conflict between Big Data interests and the respect for fundamental rights and values has provoked an increasing awareness of the dangers entailed by Big Data. Besides the above-mentioned Council of Europe initiatives, though more focused on the protection of private life, strong attention must be paid to the consequences of (mis)using depersonalised data.

People are free to decide whether they share or not share personal data for proper use. Information technology has enabled the democratisation of data, which has led to the 'Big Data' phenomenon and – at the same time – the depersonalisation of data. Once the owners loose full control over (the use and abuse of) their personal data, '**data democracy**' comes to an end, privacy is at stake and – in other words – Big Data may also threaten personal freedom. The danger of '**data dictatorship**' is that governments are tempted to misuse personal data and apply Big Data analytics in order to predict behaviour and enforce punishment, attaching more meaning to data than it deserves. Therefore, data becomes extremely important and consequently dangerous.

Massive amounts of depersonalised data are today used without consent for secondary purposes. For example, it is not unimaginable that Big Data is used for **political purposes**. Far away from traditional telephone or door-to-door polls determining the collective will of voters, since 2012 data-driven election campaigns based on 'track voting behaviour' data-mining software are successfully concluded. Without any regulation, politicians increasingly rely on data-mining analytics and profiling in order to fine-tune their political programmes and to assure reelection at the end of their mandates[22].

Finally, making money out of **trading Big Data** can be seen as a moral issue. Often lacking a staff of data scientists, (medical research, finance, retail and non-profit) companies and even governments nowadays obtain specific analyses on Big Data for their own benefits, paying data-brokers. Aside from the moral implications, this also raises the question of the (lack of) regulation of trading with data. The start of ethical debates forcing conversation between Big Data business and the public by national governments and on an international level resulting in clear data-trading rules, is most urgent, at least for the public and the data-brokers. Efficient and effective legislation regulates surveillance by personal data companies, penalises Big Data abuse committed by companies, private individuals and even officials, and ensures an effective protection of personal information.

There is a need for a Europe-wide agreement on these ethical aspects of data processing.

[22] Some countries however apply rules preventing political intrusion in private life; in Belgium, it is forbidden for politicians within a three-month period before the election date, to approach individually potential voters with gifts or commercial campaigns for voting purposes; however, this rule does not include data-mining or profiling activities.

8. Final Conclusions

Big Data analytics only work if we give organisations and governments access to personal data, sign up for Facebook or download free apps thus giving third-party companies permission to access and use our personal data, leading us to a certain point from where there is no turning back. The loss of a certain amount of privacy might be a worthwhile price to pay for all the positive innovations we will see. Problems arise though when the Big Data user deviates from the initial path for any reason whatsoever and affects the fundamental rights of providers. Big Data has enormous potential to improve society, but to reap the benefits; society must protect itself against its dark side.

GdS

Literature: "**Big Data: A Revolution That Will Transform How We Live, Work, and Think"**, by Viktor Mayer-Schönberger and Kenneth Cukier, 2013.