De Filippi, Primavera

Article

Big data, big responsibilities

Internet Policy Review

Provided in Cooperation with:
Alexander von Humboldt Institute for Internet and Society (HIIG), Berlin

Suggested Citation: De Filippi, Primavera (2014) : Big data, big responsibilities, Internet Policy Review, ISSN 2197-6775, Alexander von Humboldt Institute for Internet and Society, Berlin, Vol. 3, Iss. 1, pp. 1-12,
https://doi.org/10.14763/2014.1.227

This Version is available at:
https://hdl.handle.net/10419/213980

INTERNET POLICY REVIEW
Journal on internet regulation

Volume 3 | Issue 1

Big data, big responsibilities

Primavera De Filippi
Research and Studies Center of Administrative Science (CERSA/CNRS), Université Paris II (Panthéon-Assas), France

Published on 13 Jan 2014 | DOI: 10.14763/2014.1.227

Abstract: Big data refers to the collection and aggregation of large quantities of data produced by and about people, things or the interactions between them. With the advent of cloud computing, specialised data centres with powerful computational hardware and software resources can be used for processing and analysing a humongous amount of aggregated data coming from a variety of different sources. The analysis of such data is all the more valuable to the extent that it allows for specific patterns to be found and new correlations to be made between different datasets, so as to eventually deduce or infer new information, as well as to potentially predict behaviours or assess the likelihood for a certain event to occur. This article will focus specifically on the legal and moral obligations of online operators collecting and processing large amounts of data, to investigate the potential implications of big data analysis on the privacy of individual users and on society as a whole.

Keywords: Big data, Cloud computing, Intermediary liability, Predictive analysis, Open data, Quantified self, Privacy

Article information

Received: 17 Nov 2013 **Reviewed:** 19 Dec 2013 **Published:** 13 Jan 2014

Funding: This work is supported by the FP7 STREP project P2Pvalue - Techno-social platform for sustainable models and value generation in commons-based peer production in the Future Internet.
Competing interests: The author has declared that no competing interests exist that have influenced the text.

URL: http://policyreview.info/articles/analysis/big-data-big-responsibilities

Citation: De Filippi, P. (2014). Big data, big responsibilities. *Internet Policy Review*, *3*(1). DOI: 10.14763/2014.1.227

Big data refers to the collection and aggregation of large quantities of data produced by and about people, things or the interactions between them. These include data coming from navigation history, internet forums, social media, health records, governmental records, etc. In today's information society, as the amount of data - produced or generated - keeps growing (Swan, 2012), the aggregation and potential usages thereof are progressively affecting every facet of our life. But big data is not merely about *volume*, it is also due to the increasing *variety* of data (of different format, nature, or source) and the growing *velocity* at which it is produced and transferred through the network - a model defined by Gartner's analyst Doug Laney as the 3V's of big data.

With the advent of cloud computing, specialised data centres with powerful computational hardware and software resources can be used for processing and analysing a humongous amount of aggregated data coming from a variety of different sources (Lazar, 2012). The analysis of such data is all the more valuable to the extent that it allows for specific patterns to be found and new correlations to be made between different datasets, so as to eventually deduce or infer new information, as well as to potentially predict behaviours or assess the likelihood for a certain event to occur (Franks, 2012). As such, big data analysis has the potential to radically transform the way people act and interact online (Mayer-Schönberger & Cukier, 2013) with possible positive and negative impacts on society (Bollier & Firestone, 2010). This article will focus specifically on the legal and moral obligations of private online operators collecting and processing large amounts of data, to investigate the potential implications of big data analysis on the privacy of individual users and on society as a whole.

BIG DATA AND PERSONALISED SERVICES

Significant benefits can be derived from big data analysis. At the macro-level, big data analysis can provide valuable, and often accurate information about general economic or societal trends (Lohr, 2012). On a more individual level, processing large amounts of data provided - either willingly or unwillingly - by internet users can help online operators better understand the preferences and behaviours of their user base (La Valle & al., 2011). The identification of specific patterns of behaviour could also be used to derive precise and specific information about users, such as their current mood and state of mind, as well as their routine behaviour or current affairs (Bughin & al., 2011). From a commercial perspective, this can be very advantageous to online service providers, to the extent that it helps them provide more personalised services and better-targeted advertising (Berry & Linoff, 2004).

Yet, the question arises as to the degree to which these practices might infringe upon the privacy of end-users (Boyd & Crawford, 2012).

In Europe, the collection and processing of personal data or information is currently regulated by Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data (Data Protection Directive)[1] and Directive 2002/58/EC, as amended by Directive 2009/136/EC on privacy and electronic communications (ePrivacy Directive), which focuses more specifically on the processing of personal data in the electronic communications sector.

Article 7 of the Data Protection directive establishes the principle of *opt-in*, according to which personal data cannot be legitimately processed without the consent of the data subject, except if necessary to preserve public order or morality, as well as to further the general interest of society or individuals. Building upon this principle, Article 5 of the ePrivacy Directive further elaborates that the processing of personal data can only be achieved with the consent of the data subject - who should be given clear and comprehensive information as regard the manner and purpose of data processing - unless such processing is directly instrumental to the provision of a service which had been explicitly requested by the subject[2].

DISCLOSING PERSONAL DATA

In spite of these rules, more and more users are eager to share or disclose lots of their personal data online (Wolf, 2010), by either disclosing it publicly on the internet, or by disseminating it within a more restrained community.

The trend towards increasing data disclosure has begun with the advent of the "open data movement," which has demonstrated the benefits that could be derived from the public disclosure of public sector information (Davies, 2010). Several tools for data analysis and visualisation have been developed to enhance the legibility and comprehensibility of such data, so as to ultimately promote more transparency and accountability in the public sector. These tools have been rapidly adopted by the private sector, which primarily deployed them for commercial purposes, with a view to enhance their marketing strategies based on the analysis of customer preferences and behavioural data. In this regard, big data analysis has shown that significant value can be extracted from the collection and aggregation of data into large datasets so as to be able to better identify patterns and correlations amongst them (Lohr, 2012). Today, with the emerging quantified self (QS) movement, most of the attention is shifting towards personal data. As more and more people benefit from sharing or exchanging personal information with the world around them, the public disclosure of personal information is, indeed, progressively turning into a trend (Swan, 2013). The recent growth in popularity of activity trackers (such as the Fitbit, Basis, or Nike's Fuelband) clearly illustrates that users are becoming increasingly comfortable with disclosing personal information (including health data) in exchange of highly personalised services tailored to their own preferences and needs. Yet, with these devices, users are not just collecting data about themselves; they rely on third party operators to process a large amount of personal information coming from many different sources (Richards & King, 2013), and to analyse it by means of statistics and sophisticated data analysis techniques in order to extract new information that could not be easily inferred from the individual dataset of any single person (Cambria & al., 2013).

While this is likely to impinge upon the individual right to privacy and data protection (Craig & Ludloff, 2012), in the case of many online services, consent is obtained by means of long and intricate Terms of Service (ToS) which users necessarily have to agree to in order to benefit from the services offered by online operators (Bradshaw & al., 2011). Most users do, in fact, generally agree to their personal data being collected and processed by third party online operators in order to benefit from a more customised service that would not be possible otherwise (Oboler, 2012).

Yet, privacy policies are often way too complicate to be properly (or fully) understood, and are therefore, oftentimes, simply ignored by users who might end-up giving uninformed consent to the processing of their personal data (Thompson, 2012).

Besides, even where users have consciously consented to these terms, by the mere fact of aggregating different datasets together, online operators can deduce or infer new information about their user-base, which often goes beyond the information that was explicitly or implicitly provided by them (Witten & Frank, 2005).

In this respect, extensive data analysis based on the aggregation of multiple datasets coming from a variety of different sources raises a series of concerns as regards the right to anonymity or pseudonymity of certain users who might rather maintain separate identities offline and online, or who might even impersonate different online identities according to the communities

they interact with (Bus & Nguyen, 2013). While it is often easy for online operators to find a correlation between different user profiles (e.g., because they share a common attribute such as their email or phone number), users might be unwilling to disseminate the information disclosed (either publicly or privately) by one of their identities beyond the community of people with whom that profile was actually meant to interact.

This is particularly relevant in the context of large datasets, which have been purposely 'anonymised' in order to comply with privacy and data protection regulations. Indeed, if the law only applies to the processing of personal data – *i.e.*, data related to individuals who *are* or *can* be identified, either directly or indirectly, by virtue of one or more of their distinctive characteristics [3] - it also stipulates that, in order to establish whether or not a person can actually be identified, it is necessary to take into account all means available to the public for achieving such identification [4]. Thus, in the case of anonymised datasets, although the identity of users is effectively protected when every dataset is taken independently, certain individuals could nonetheless be re-identified by aggregating data coming from multiple data sources into one large dataset so as to find new patterns and correlation – a so-called 'inference attack' [5].

Thus, as the amount of data collected keeps growing, it becomes more and more likely that large online operators, such as Google, Facebook, or Amazon eventually become much more aware of their users' interests, preferences and personal information[6] than users might think (Cumbley & Church, 2013).

This raises, again, a series of privacy concerns to the extent that – even if users did actually consent to the processing of some of their personal data - they did not explicitly consent to the collection (or, in this case, the extrapolation) and processing of information which has been derived from it by means of big data analysis (Kerr & Earle, 2013). Insofar as such information qualifies as personal data (because it relates to the preferences or distinctive characteristics of a person), online operators should, at least theoretically, be unable to benefit from the processing of such data without the prior consent of the data subject.

In this respect, notice constitutes an essential precondition for informed consent. Yet, the traditional model of notice and consent - based on a precise definition of the type of data that will be collected, together with a specific delineation of the purpose for which such data will be processed[7] - does not, however, adequately match the transformative nature of big data, which is constantly growing and evolving over time (Cukier & Mayer-Schoenberger, 2013). Indeed, in the context of big data, the data provided, either directly or indirectly, by users constitute the basis from which to deduce or infer new (personal) data. Much of the value that can be extracted from big data analysis is, however, not perceptible at the time of collection, which is generally when notice and consent are given. Users are often not even aware of the fact that more data is being produced about them, and cannot therefore be expected to consent to the processing thereof. Nevertheless, given that most of the data processing is performed locally on-premises, and that the outcome is not made publicly available, but rather used internally for the mere purpose of providing a more customised and personalised service, it is difficult for users to realise that their right to privacy is being effectively infringed upon.

More critically, in spite of the growing sophistication of statistical tools and inferential algorithms, the accuracy of the resulting information can obviously be put into question (Kaisler, 2013). In this regard, not only does the Data Protection Directive establish a right for citizens to gain access to all of their personal data held by a third party, as well as to be informed of the actual usage thereof (Article 12) but it also gives them the opportunity to object to such usage (Article 14), as well as to correct the data that they consider to be inaccurate (Article 12).

Yet, given that big data analysis is generally carried out internally, in order to provide personalised advertising or better service customisation, users are often not properly notified of the fact that online operators have inferred additional information (related e.g., to their personal preferences, habits, family status, medical history, or financial situation). Hence, to the extent that they are not even aware of such inferences, they cannot readily oppose to the processing thereof, nor can they submit a request for that information to be rectified, if necessary (Cavoukian & Jonas, 2012).

BIG DATA AND PUBLIC ORDER

Beyond personal recommendations systems or sophisticated mechanisms of customisation aimed at providing a more personalised service to users, online service providers might potentially rely on big data analysis to deduce or infer information that could potentially be used to further the general interest of society.

Big data analysis can also be used to identify certain patterns of behaviour and typologies of users (Brown & al., 2011). As the amount of data increases, the accuracy and reliability of these analyses also increase. By analysing data from a variety of sources, specific algorithms have been developed to predict certain behaviours (or misbehaviours), as well as to infer information about the private life of individuals (Truvé, 2011).

For instance, by relying on advanced data-mining techniques and statistical correlation algorithms, credit card companies can identify customers who are having a love affair, recognise those who recently moved into a new house, or even predict an imminent divorce. Similarly, by combining a patient's health records with the data collected by wearable devices or personal sensors, health-tracking services can identify underlying disorders and protect their patient from imminent diseases (Barrett & al., 2013). More generally, by analysing individual internet usage patterns (combining search and navigation history, with the schedule and speed of navigation, etc), it is possible to establish the mood and personality of users, determine their current state of mind, or even identify specific signs of depression.

The question arises, therefore, as to whether - given their exclusive access to a huge volume of information about users - large online operators are under the moral responsibility (Han, 2013) to intervene *ex-ante* in order to promote or reprimand certain types of behaviours.

Both in Europe (see e.g., Articles 3(2) and 7(e) of the European Data Protection Directive; the Regulation of Investigatory Powers Act of 2000 in the UK, and the recently enacted 2014-2019 Defense Bill in France) and in the U.S. (see e.g., the PATRIOT Act and the Foreign Intelligence Surveillance Act), the law stipulates that, for the purposes of national security and for preserving the legitimate interests of society, the government has the right to request online operators to disclose personal data and information related to individuals suspected to be involved in criminal activities. But does the reverse hold true? Do online operators have a duty or moral obligation to disclose the information they have gathered about their user-base - through data mining techniques and inferential analyses – in order to protect individual users from an imminent danger or, more generally, to promote the general interest of society?

For instance, if a health-tracking service operator believes that one of his users is under the risk of incurring diabetes, does he have the right to communicate such information, along with that user's daily nutritional habits, to a doctor? If online operators are able to identify users who

have fallen into depression and whose patterns of behaviour indicate that they might be considering to commit suicide, do they have the right (or duty) to protect these users by preventing access to certain pieces of content and information and/or by communicating these symptoms to a public health agency?

While the answer to these questions is, ultimately, a matter of degree, the issue becomes more critical when the outcome of user profiling and data analysis reveals a user's tendency to violent behaviours, or a propensity to engage into criminal activities.

Government agencies are already employing advanced data mining techniques in order to identify potential evidences of terrorist activities (Mena, 2003). Surveillance cameras are increasingly deployed along with analytical software designed to detect specific patterns of behaviour that might require the intervention of the police force. More generally, statistical and analytical tools can be used to assess the likelihood for a crime to be committed in a certain area or place.

While this might be regarded (by some) as a legitimate activity when performed by public authorities acting within their mission to preserve public order and morality, how does this actually translate to the private sector? Do online operators - like Google, Twitter or Facebook - have the right to impinge upon the user's right to privacy by reporting suspicious activities to the police, or suggesting that users exhibiting the behavioural pattern of a criminal be put under surveillance?

MAIN PRIVACY RISKS

1. Anonymisation and re-identification: No dataset can be perfectly anonymised. By drawing a series of correlations between different datasets, one can obtain information that could potentially contribute to identifying one or more data subjects (a so-called "inference attack").

2. Right to be forgotten: While the law stipulates that users have the right to request information about themselves to be deleted or rectified, this does not necessarily apply to the information inferred as a result of big data analysis.

3. User discrimination: User profiling can lead to discrimination not only according to the personal data provided by users (e.g., age,

gender, ethnic background) but also according to the information inferred from big data analysis (e.g., health condition, social background, etc).

4. Pattern-matching: Big data analysis and pattern matching techniques sometimes produces result that could be mistaken for reality. Users are profiled into specific categories on the basis of their past and current behaviours, regardless of whether they have actually proven or confirmed to belong into one of these categories.

5. User profiling: Although user profiles are derived from their former behaviours, prescriptive analysis also constitutes a driver for future behaviour, creating a loop that will ultimately reinforce the profile which has been assigned to each user, and thereby turning a prediction into reality.

The issue is highly controversial. Although this might lead to a safer and more controlled society, there are, however, great risks in endowing online operators with the right (or even the duty) to disclose personal information to public and/or private authorities, merely on the basis of users' past and current behaviours.

Even though the legitimacy of such disclosure is, nowadays, still unclear, in Europe, Article 15 of the Data Protection Directive stipulates that, while it might be considered in court, inferred information resulting from automated processing and profiling cannot constitute the *sole basis* for a court order [8].

Yet, regardless of whether or not personal data or information can actually be disclosed to the authorities, the question remains as to whether online operators have a moral obligation to act upon foreseeing a crime – *i.e.*, do they have a duty to constrain users' behaviour in order to reduce the potential for illegal activities?

Again, this is a controversial issue. While it might indeed reduce the number of crimes committed online (and offline), acting upon statistical analyses in order to prevent harmful or objectionable activities from happening, on the mere basis of predicted behaviour, might actually lead to a series of unjust limitations on the freedom of certain users, who are *presumed*

guilty before proven innocent. If it suffices to reflect a certain pattern of behaviour in order to become a 'suspect', the constitutional presumption of innocence might, indeed, be put into jeopardy (Kerr & Earle, 2013).

Something that heavily resembled science fiction just a few years ago (see e.g., George Orwell's "1984", Philip K. Dick's "Minority Report" or Andrew Niccol's "Gattaca") is now increasingly becoming part of reality. Today, while online operators do not (yet) prevent us from acting in one way or another according to our past behaviour, they are, however, to a certain level already limiting our ability to freely choose the content or information that will be offered to us, according to sophisticated algorithms aggregating our profiles, attitudes and activities, and processing them into preferences (Bollier & Firestone, 2010). Depending upon the category they fall into, different users will thus be subject to a different collection of content and information, or provided with a different typology of services that might significantly affect much of their online practices (Pazzani & Billsus, 2007). As such, recommendations systems - albeit useful and appreciated by many (Lops & al., 2011) - represent the first step towards the establishment of a condescending system, where users' behaviours are increasingly guided - and, to a large extent, determined - by their own data and profiles.

Thus, in the age of cloud computing, where almost everything we do online can be tracked by a variety of internet service providers (Stein, 2011), it becomes ever more important to acknowledge that the benefits derived from big data analyses have to be counterweighted with the risks resulting from third party online operators collecting (and inferring) too much information about us (Bollier & Firestone, 2010). On the one hand, user-profiling by aggregating data from many different data sources enable users to benefit from a service that is more valuable to them insofar as it is more in line with their individual preferences or tastes (Brown & al., 2011). On the other hand, however, extensive data collection and analysis allows for online operators to acquire in-depth knowledge about their user-base, which might subsequently be used to assess current activities or predict the future behaviour of users (McAfee & Brynjolfsson, 2012). Provided these predictions can actually be acted upon, online operators have the ability to substantially influence user behaviour: rather than being the consequences of former online practices, prescriptive analysis might eventually shape the future behaviours of users - thereby incurring the risk of turning a prediction into reality.

**The original version of this article published Jan 13, 2014 has been revised. According to Golle (2006) gender, birthday and ZIP code are enough to uniquely identify over 63% of the U.S. population from publicly available databases, not 87 % as stated before.*

FOOTNOTES

1. The Data Protection Directive of 1995 will eventually be replaced by the proposed General Data Protection Regulation (GDPR) whose goal is to harmonise data protection regulations throughout the EU, while taking into account important aspects like globalisation and new technological developments (such as cloud computing and social networks) that were not properly accounted for in the directive.

2. See Article 5(3) of the ePrivacy Directive, stipulating that "the storing of information, or the gaining of access to information already stored, in the terminal equipment of a subscriber or user is only allowed on condition that the subscriber or user concerned has given his or her consent, having been provided with clear and comprehensive information [...] about the purposes of the processing. This shall not prevent any technical storage or access for the sole purpose of carrying out the transmission of a communication over an electronic

communications network, or as strictly necessary in order for the provider of an information society service explicitly requested by the subscriber or user to provide the service."

3. See article 2 of the Data Protection Directive, stipulating that "'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;"

4. See "whereas 26" of the Data Protection Directive, stipulating that "the principles of protection must apply to any information concerning an identified or identifiable person; whereas, to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person."

5. According to a study by Golle (2006), gender, birthday and ZIP code are enough to uniquely identify over 63%* of the U.S. population from publicly available databases.

6. See e.g., in the financial sector, banks increasingly relying on big data analysis in order to determine whether or not to issue a loan or a mortgage to individuals or companies, in the medical field, the case of research being undertaken on data provided for one purpose and being ultimately employed for alternative purposes; or the case of various supermarkets, which can predict personal information about their customers (such as pregnancy for instance) by monitoring their purchasing patterns and behaviours.

7. See Art. 6(1)(b) of the EU Directive 95/46/EC, which stipulates that personal data must be "collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes."

8. See article 15 of the Data Protection Directive, stipulating that "Member States shall grant the right to every person not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc."

REFERENCES

Arthur, L. (2013). *Big Data Marketing: Engage Your Customers More Effectively and Drive Value*. John Wiley & Sons.

Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. Wiley. com.

Barrett, M. A., Humblet, O., Hiatt, R. A., & Adler, N. E. (2013). Big Data and Disease Prevention: From Quantified Self to Quantified Communities. *Big Data*, 1(3), 168-175

Bollier, D., & Firestone, C. M. (2010). *The promise and peril of big data* (p. 56). Washington, DC, USA: Aspen Institute, Communications and Society Program

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679.

Bradshaw, S., Millard, C., & Walden, I. (2011). Contracts for clouds: comparison and analysis of the Terms and Conditions of cloud computing services. *International Journal of Law and Information Technology*, 19(3), 187-223.

Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'?. *McKinsey Quarterly*, 4, 24-35.

Bughin, J., Livingston, J., & Marwaha, S. (2011). Seizing the potential of 'big data'. *McKinsey Quarterly*, 103-109.

Bus, J., & Nguyen, M. H. C. (2013). Personal Data Management–A Structured Discussion. *Digital Enlightenment Yearbook 2013: The Value of Personal Data*, 270.

Cavoukian, A., & Jonas, J. (2012). Privacy by design in the age of big data. *Office of the Information and Privacy Commissioner*.

Cambria, E., Rajagopal, D., Olsher, D., & Das, D. (2013). Big social data analysis. *Big Data Computing*, 401-414.

Craig, T., & Ludloff, M. E. (2011). *Privacy and big data*. O'Reilly Media, Inc.

Cukier, K., & Mayer-Schoenberger, V. (2013). Rise of Big Data: How it's Changing the Way We Think about the World, The. *Foreign Affairs*. May-June Issue.

Cumbley, R., & Church, P. (2013). Is "Big Data" creepy?. *Computer Law & Security Review*, 29(5), 601-609.

Davies, T. (2010). *Open data, democracy and public sector reform: A look at open government data use from data. gov. uk*. Practical Participation.

Franks, B. (2012). *Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics* (Vol. 56). Wiley. com.

Golle, P. (2006). Revisiting the uniqueness of simple demographics in the US population. In *Proceedings of the 5th ACM workshop on Privacy in electronic society* (pp. 77-80). ACM.

Han, J. (2013). The ethics of big data. *Living Ethics: Newsletter of the St. James Ethics Centre* (92), 7.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January). Big Data: Issues and Challenges Moving Forward. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp. 995-1004). IEEE.

Kerr, I., & Earle, J. (2013). Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy. *Stanford Law Review Online*, 66, 65.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, 52(2), 21-31.

Lazar, N. (2012). The Big Picture: Big Data Hits the Big Time. *CHANCE*, 25(3), 47-49.

Lessig, L. (1999). *Code: And other laws of cyberspace*. Basic Books (AZ).

Lohr, S. (2012). The age of big data. *New York Times*, 11.

Lops, P., de Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In *Recommender Systems Handbook* (pp. 73-105). Springer US.

Mayer-Schönberger, V., & Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live. *Work and Think*. London: John Murray

McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 60-66.

Mena, J. (2003). *Investigative data mining for security and criminal detection*. Butterworth-Heinemann.

Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web* (pp. 325-341). Springer Berlin Heidelberg.

Richards, N. M., & King, J. H. (2013). Three Paradoxes of Big Data. *Stanford Law Review Online*, 66, 41.

Stein, J. (2011). *Data mining: How companies now know everything about you*. Time Magazine.

Swan, M. (2012). *Sensor mania! The Internet of Things, wearable computing, objective metrics, and the Quantified Self*. 2.0. Journal of Sensor and Actuator Networks, 1(3), 217-253.

Swan, M. (2013). The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. *Big Data*, 1(2), 85-99.

Oboler, A., Welsh, K., & Cruz, L. (2012). The danger of big data: Social media as computational social science. *First Monday*, 17(7).

Thompson, D. (2012). *"I Agreed to What?": A Call for Enforcement of Clarity in the Presentation of Privacy Policies*. Hastings Comm. & Ent. LJ, 35, 199-223.

Truvé, S. (2011). *Big Data for the future: Unlocking the predictive power of the Web*. Recorded

Future, Cambridge, MA, Tech. Rep.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Wolf, G. (2010). The data-driven life. *New York Times*, 28.