

# Big Data Analytics: An Ethical Question

## Research Essay

Leah Korganowski  
3108793  
COMP 607 – Fall 2015

**Abstract**—Data tracking and collection has never been easier and more detailed than it is today. Corporations including media, advertisement, communications, commerce and communications companies are all taking advantage of the data trails that users leave behind in their everyday lives. However many consumers are unaware that this phenomena is happening and that potentially identifying and private information is being analyzed and sold to other third parties that they have no affiliation with. Ethics is raised in the context of what the line is between normal business behaviour and infringing on human rights of privacy and dignity. Such companies as Verizon, Netflix, and Google to name a few all participate in some form of data collection. Many of which use the data collected to profit off of those consumers. Major questions that society needs to raise is whether this is ethical and how to rectify the situation. Such approaches as: knowing the data source, classifying the data, transparency and accessibility, consent, and education all can play a role to ensure companies uphold a high ethical standard when it comes to using data.

### I. INTRODUCTION

With the ever increasing presence of social media and mobile use companies have been taking advantage of the increasing amounts of data trails to create market targets and predictions. This collection of data is now raising concerns over the ethical conduct of how the data is being collected, stored, and shared, as it raises a number of red flags. Privacy infringements have already been committed by some of the largest companies in the world. Even though there are laws governing privacy, they do not transcend jurisdictions and often have many loopholes that data companies are able to bypass. To ensure that data companies are complying with a basic ethical standard there needs to be standardized approaches to handling data.

### II. AN ETHICAL APPROACH

Traditionally ethics were based on a more normative approach, which involved following the right and wrong behaviour expectations based on a societies moral standards [1]. But how does ethics play a role when the context and the detail changes? Such questions are raised in regards to Big Data, which spans globally and across cultures.

Another important concept to consider in terms of the use analytics and ethics is harm. As mentioned above in normative ethics harm is more easily determinable, but in terms of privacy how is this measured if there is no physical harm occurring? In order to determine this a variety of active ethical theories must be looked at to determine what constitutes ethics as well as look at ethical problem solving. It is

then important that we as a society and corporations take responsibility and follow an ethical standard.

Velasquez et. al details on five main approaches to ethics that are used today [2]. To briefly review they include: Utilitarian; determines that the most ethical action should be the one that provides the largest good for the largest number, Rights; this approach is more open ended but includes that humans have dignity and have the right to truth, privacy, etc., Fairness of Justice; sees favouritism and discrimination as not right and that individuals should be treated equally, Common-Good; social policies, systems, and environments are good to all, Virtue; individuals should strive towards certain ideals[2]. With these five main ethical theories, Velasquez et. al suggests that we can use them to come to a general approach to situations where ethics are of concern [2]. This can then be applied to big data and its ethical implications by corporations. Such important questions that can then be asked are:

1. *What are the benefits and harm that big data produces? What actions can lead to the best overall consequence?*
2. *What are the moral rights of the affected parties? What actions should companies be playing to respect these rights?*
3. *Is there a usage of big data that treats everyone the same or is there discrimination?*
4. *What course can companies dealing in big data take to benefit the common good?*
5. *Can these companies take an action which develops and enhancing moral virtues?*

[2]

### III. BIG DATA – THE CONTEXT

Big Data is a relatively new term, coming into being over the last decade with the massive use of the internet and social media. In essence it can be described as a large amount of data which can be obtained from a variety of sources, such as social media, telecommunications data, and search engine information [3]. This information is of importance to many large companies because they are able to take the data and measure and determine insights into their firms to improve business and marketing decisions [4]. There are a variety of applications that Big Data is used in, some of the more common ones including: determining any issues in real time, determine trends of buying behaviour and market to target consumers, determine customer buying habits, target fraudulent behaviour and risk, and many other uses [3].

#### A. Why It Is Used

Firms use big data for a number of primary reasons. Them being:

- Cost reduction
- Faster and accurate decision making
- Predictive analytics
- Create new products and services
- Surveillance and consumer behaviour tracking
- Purchasing Behaviour tracking
- Ability to create real-time custom tailored advertisements

[5] [6]

Big data can help with predictive analytics for companies, which predict where trends or services will be. Which includes taking customer data from viewed shows, time spent watching and forecasting customer trends, looking at behavioural patterns of watching, and creating and marketing certain products based on those trends and behaviour.

#### B. Big Data in Use – Case Studies

Wireless operators are a primary user for big data, as they have access to millions of data trails perfect for analytics. In particular Verizon is a big user of this data and in 2011 it actively changed its privacy policy to share customer data with parties outside of the company [7]. This includes selling its information about its mobile phone customers, which gives away locations, activities, backgrounds and almost all other data trails possible [3]. The story that this information can tell is outstanding, from these data trails they can determine demographics and app interests [7]. These actions are some of the many that seem questionable in the eyes of individual privacy and rights. To what extent can something as personal as our personal

cellphone, which many consider as another limb, to be actually a tracking and profit mechanism for profit-based and advertisement companies?

Netflix also uses predictive analytics which is sourced from Big Data to create new shows and to market them to their consumers based on their behaviour patterns [6]. This further demonstrates that these analytical techniques are not in isolation, but rather common place for many service providers.

However there are some companies that are taking responsibility and action for big data use. Verizon as mentioned above, is known originally for selling its user data to outside parties. Recently they just started up a pilot effort to allow customers to opt in to sharing their cellphone data, rather than opting out [7]. It must be noted though that this is a pilot effort. The question to be raised is how the company will approach this when the majority of customers do not opt in. Users should be aware that they do profit from selling data and gain a competitive advantage, therefore they may not let this go in the name of respecting individual privacy?

Another company to look at is Google. Google is considered to be one of the major users of Big Data, and their usage of the data can be seen directly back to the user of their search engine in tailored advertisements and suggested search options. The approach to collecting and analyzing individual data is quite simple to Google, when a word or phrase is typed into the search bar Google then analyzes the word in both literal and semantic searches. [8]



Fig. 1. Google Semantics. [8]

In the literal search, Google looks find exact words or phrases to return with. In the semantic search it also attempts to understand the search context (Fig. 1.). The interesting component of this is not just the search result the user is left with, but what is happening in the backend of Google’s system. If users are logged into their Google account, their previous history, location, and keywords are used by Google. This

history is then added to their keyword synonyms and search engine’s syntactic algorithm to predict product relationships of interest to them. The site essentially isn’t just providing a literal search engine service, which is what users use it for, but rather suggesting products and advertisement of relation that the user may want to buy. The user is thus essentially throwing themselves to marketers. What is questionable about Google analytics is when one’s search topics extend to other sites, such as Facebook, YouTube, and other sites, with advertisements related to one’s previous searches. [8]

#### IV. MISUSES OF DATA

While Big Data definitely has benefits, it is primarily the corporations and data analytic companies that have any gain. But what is the line that should be drawn when it comes to data acquisition and analytics? Are companies using and selling consumer data at a right or has this gone too far? These are all the questions that ethics takes into consideration. There have been a variety of data misuses, some which include over convoluted subscriber policies, data used in harmful ways, and the emerging distrust between consumer and corporations.

##### C. Data Policies

While there is governing legislation on privacy laws and reporting, an issue is that often they do not transcend borders and are not written well enough to encompass all aspects of data use. One example as pointed out by Carson (2015), is in regards to The Fair Credit Reporting Act (FCRA), which protects individuals from data being provided to parties. However in regards to this act it does not apply to groups, allowing user data to be used and shared when it is in bulk form [9].

Such are policies about collecting low sensitivity data, which when combined and analyzed in a certain way can be used maliciously and in reidentification to identify the data’s original user. The example that Buytendijk and Heiser (2013) give is an individual who has knowingly given permission to give their data as anonymous, not knowing that reidentification is able to be done [10]. They also state an interesting fact that 87% of the US’s citizens that are in public databases can be identified using three anonymous data points: birthday, gender and postal code. This is known as the triple identifier [10].

##### D. Data Brokers

Data brokers are a primary concern in the Big Data field. They are the companies that collect and manage consumer data and for a price sell it to other

companies. The issue with these companies is their methods. Riglian (2012) identified that in many circumstances the web browser option for do not track for consumers is often ignored by these companies, as it is not legally binding, allowing for these companies to ignore ones opt out and thus track the consumer’s data [11].

The Federal Trade Commission (FTC) identified formally that they have discovered that the data broker industry does not operate with transparency and that data that is collected through these companies is shifted across many entities [9]. There is a risk associated with the more hands the data crosses, such as whether or not strict confidentiality agreements upheld and enforced in a consistent manner across all data participants. In order to ensure that the original user’s data is being protected there needs to be a governing law against data broker misconduct.

##### E. Creating a distrustful world

With all the secrecy and transferring of information, big data can have a negative long term consequence for companies and consumers. The abuse and misuse of personal information could create a distrust between individuals and the applications that they use. This is important to mention because as a relatively new phenomena, Big Data is still in its infancy and the long term effects are still widely unknown.

#### V. CORPORATE ETHICS IMPORTANCE

Ethics are important to demonstrate, as they instill trust, integrity and risk reduction to firms [12]. In order for big data corporations and data brokers to conduct business in an ethical matter they should implement the following strategies:

Ethical Approach	Description
Be aware of the data source	<ul style="list-style-type: none"> <li>Know and keep track of where data is coming from</li> <li>Have a regulated screening process that ensure the data collected was through legitimate means, via notice and choice. [9]</li> </ul>
Classify the data	<ul style="list-style-type: none"> <li>Don’t just keep all data in one bucket of accessibility.</li> <li>Separate sensitive and financial information from innocuous data</li> <li>Set higher standards of care and security for the more sensitive data</li> </ul>

	[9]
Transparency and Accessibility	<ul style="list-style-type: none"> <li>Keep records of data keeping and collection methods</li> <li>Show consumers their data. The majority of consumers don't opt out of their data collection even though many companies allow them to.</li> </ul> [9]
Consent	<ul style="list-style-type: none"> <li>Consent should be required by law and enforced on any companies that do not comply.</li> <li>It should be presented in a way that is understandable to the average person (not use legal jargon)</li> </ul>
Education	<p>Education in terms of ethical obligation in big data involves two important actions:</p> <ol style="list-style-type: none"> <li>1) Companies dealing with the data need to define ethics within the organization and follow it.</li> <li>2) Public must become data conscious aware citizens.</li> </ol> [9]

Fig. 2. Ethical Approaches in Big Data.

These ethical approaches are key to ensuring that big data organizations are following a standardized method of obtaining, managing and sharing data. It is obvious that there are lines being crossed in terms of ethical conduct with data present day.

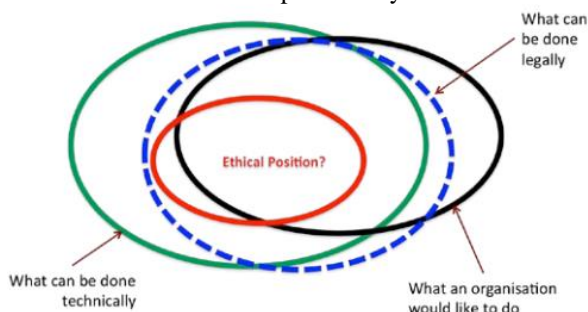


Fig. 3. Ethics Position in Big Data. [13]

As demonstrated in Fig. 3. ethics has a unique position within the big data industry, allowing companies to fall under the class of socially responsible if they balance between the law, technicalities, and the wants of the organization.

## I. CONCLUSION

Ethics is an obvious issue in terms of corporate responsibility and managing user data. Although there are many benefits to the collection, distribution, and analysis of big data, it also poses threats to individual trust and the essential basic freedom of privacy rights. The area of ethics in this industry is hard to define as big data often spans across borders, cultures, and jurisdictions. However based on common ethical approaches, including transparency, accessibility, consent, and education, corporations would be able to follow a general and consistent means to looking out for their user's best interests. In order to make this come into being though there needs to be a governing organization over big data use that enforces legislation to protect the rights of user data online.

## REFERENCES

- [1] Fieser, J. "Ethics." Internet: <http://www.iep.utm.edu/ethics/#H2>, 2015 [Dec. 8, 2015].
- [2] Valasquez, M. et al. "Thinking Ethically: A Framework for Moral Decision Making." Internet: <http://www.scu.edu/ethics/practicing/decision/thinking.html>, 1996 [Dec. 8, 2015].
- [3] Davenport, T. "Three big benefits of big data analytics." Internet: [http://www.sas.com/en\\_sa/news/sascom/2014q3/Big-data-davenport.html](http://www.sas.com/en_sa/news/sascom/2014q3/Big-data-davenport.html), 2014 [Dec. 8, 2015].
- [4] McAfee, A. "Big Data: The Management Revolution." Internet: <https://hbr.org/2012/10/big-data-the-management-revolution/ar>, 2012 [Dec. 8, 2015].
- [5] Chouffani, R. "Big Data Analytics a Big Benefit for Marketing Departments." Internet: <http://www.cio.com/article/2392067/data-management/big-data-analytics-a-big-benefit-for-marketing-departments.html>, Sept. 24, 2012 [Dec. 8, 2015].
- [6] Lange, L. "Big Benefits from Big Data." Internet: [http://www.sap.com/bin/sapcom/lt\\_lt/downloadasset.2014-09-sep-11-15.big-benefits-from-big-data-pdf.html](http://www.sap.com/bin/sapcom/lt_lt/downloadasset.2014-09-sep-11-15.big-benefits-from-big-data-pdf.html), 2014 [Dec. 9, 2015].
- [7] Leber, J. "How Wireless Carriers Are Monetizing Your Movements." Internet: <http://www.technologyreview.com/news/513016/how-wireless-carriers-are-monetizing-your-movements/>, Apr. 12, 2013 [Dec. 9, 2015].
- [8] Rijmenam, M. V. "How Google Applies Big Data To Know You – Infographic." Internet: <https://datafloq.com/read/google-applies-big-data-infographic/385>, 2015 [Dec. 9, 2015].
- [9] Carson, A. "Data Brokers Demystified: A Call for Ethics." Internet: <https://iapp.org/news/a/data-brokers-demystified-a-call-for-ethics/>, Sept. 18, 2014 [Dec. 10, 2015].

- [10] Buytendijk, F. and Heiser, J. "Confronting the privacy and ethical risks of Big Data." Internet: <http://www.ft.com/cms/s/0/105e30a4-2549-11e3-b349-00144feab7de.html#axzz3uF75u7Sf>, Sept. 24, 2013 [Dec. 10, 2015].
- [11] Riglian, A. "‘Big data’ collection efforts spark an information ethics debate." Internet: <http://searchcloudapplications.techtarget.com/feature/Big-data-collection-efforts-spark-an-information-ethics-debate>, [Dec. 10, 2015].
- [12] Morgan, L. "Big Data Ethics: 8 Key Facts To Ponder." Internet: [http://www.informationweek.com/big-data/big-data-analytics/big-data-ethics-8-key-facts-to-ponder/d/d-id/1322143?image\\_number=9](http://www.informationweek.com/big-data/big-data-analytics/big-data-ethics-8-key-facts-to-ponder/d/d-id/1322143?image_number=9), Sept. 25, 2015 [Dec. 10, 2015].
- [13] Chessell, M. "Ethics for big data and analytics." Internet: [http://www.ibmbigdatahub.com/sites/default/files/whitepapers\\_reports\\_file/TCG%20Study%20Report%20-%20Ethics%20for%20BD%26A.pdf](http://www.ibmbigdatahub.com/sites/default/files/whitepapers_reports_file/TCG%20Study%20Report%20-%20Ethics%20for%20BD%26A.pdf), 2014, [Dec. 10, 2015].
- [14] SAS. "Big Data." Internet: [http://www.sas.com/en\\_us/insights/big-data/what-is-big-data.html](http://www.sas.com/en_us/insights/big-data/what-is-big-data.html), 2015 [Dec. 10, 2015].