ESSAY COVER SHEET 2018

NIR605: Critical Data Studies

Instructions:

All work must be typed and pages must be numbered. This cover sheet must be clearly visible at the front of all work.

Full Name: **Sean O'Riogain**

Student Number: **18145426**

Date Submitted: **6th May 2019**

Course Title & Lecturer: **Critical Data Studies (NIR605), Dr. Rob Kitchin.**

Essay/Project title: **A Critical Examination of the Ethics of Big Data and the Data Brokerage Industry**

NIR605 – Critical Data Studies - Assignment 3

A Critical Examination of the Ethics of Big Data and the Data Brokerage Industry

Introduction

The aim of this paper is to explore the ethical challenges that the big data phenomenon, and the data brokerage industry that it has spawned, pose to modern society and how those challenges might be met. Before doing so, it seeks to explain what data brokers are and how they came into being and developed. In so doing, the intention is that those ethical issues will be exposed and that, by contextualising them and the relevant actors, a new approach to resolving them can be identified.

Setting the Scene

With the advent of the era of big data, as examined in the detail in the book *The Data Revolution* (Kitchen, 2014), a huge and exponentially-growing amount of data on almost every aspect of human endeavour is being continuously captured in real time (and otherwise) in a variety of digital formats (Laney, 2001).

The enablers of this phenomenon are primarily of a technical nature ranging from the invention of the internet and, especially, its worldwide web manifestation, the ever-increasing level of processing power that computing technology is able to achieve, including its capacity to store, retrieve, manage, process and analyse those data, to the advent of, and advances in, Information Communications Technologies (ICTs) such as the smartphone. This, in turn, has led to the interconnection of vast numbers (now in the billions) of devices - from computers, mobile devices, cameras and sensors (e.g. in vehicles, domestic appliances, medical devices, fitness trackers etc.) - a phenomenon that is known as the Internet of Things (IoT). The advent of worldwide web has also led to the development of social media applications, like Facebook, Twitter, Instagram and YouTube, which also contribute a huge amount of unstructured data (e.g. free-format text, images, video etc.) to the available pool of big data. The arrival of financial products such as credit and debit cards, online banking and mobile payment applications such as PayPal, Google Pay and Apple Pay have facilitated the growth of huge online retailers like Amazon which, in turn, have forced traditional 'bricks-and-mortar' retailers to deploy significant online sales and marketing channels of their own.

All of this means that in the western world every aspect of human life is capable of being constantly recorded in digital form depending on an individual's use – either consciously or unconsciously – of the aforementioned types of technology. The data that is directly generated by a person through use is commonly known as his or her digital footprint while the data that is generated about them is called his or her digital (or data) shadow (Kitchin, 2014). Entities in both the public and, particularly, the private sectors are the main drivers of the deployment of those technologies and for the harvesting and harnessing of the resultant data flows. In parallel, the public sector is doing the same thing for the 'small', but important, data that track an individual's progress through life from the cradle to the grave including the digitisation of their historical equivalents. This type of (often publicly-available) data includes birth, marriage, divorce and death records as well as details of an individual's educational attainment, drivers licence, employment details (including earnings for tax purposes), social security details and welfare payments, dependents and child support, planning applications, property acquisitions and disposals (again for taxation purposes mainly), driving license, military service, voter registration and criminal records etc., etc.

Where Does the Data Brokerage Industry Come In?

When captured correctly, kept up to date and conjoined effectively, the data from all of the sources referred to above (as well as many others) can provide a comprehensive view of an individual's movements, achievements, resources (financial and otherwise), capabilities, needs, wants, interests, beliefs, and cultural and political allegiances.

Of course, all of this information is potentially an extremely valuable resource for a huge variety of entities in both the public sector (e.g. law enforcement, state security, immigration, taxation and social welfare authorities, political campaigners etc.) and private sectors (marketers and sellers of goods and services, recruiters etc.).

And that is where the data brokerage industry comes in – to perform the type of data capture, curation and aggregation referred to in the paragraph before last. In some shape or form, data brokers have existed for decades. In the early days they offered mailing lists to advertising and marketing companies (and still do). With the advent of the big data phenomenon they have expanded hugely in terms of their numbers, the volumes of data that they acquire, the types of analysis they offer, the numbers and types of businesses that they serve and the level of economic activity and wealth that they generate for themselves and others.

In the US, the major participants in the data brokerage market include 411.info, CoreLogic, DataLogix, eBureau, Experion, Epsilon, FICO, Harte Hanks, infoUSA, Instant Checkmate, Intelius, LexisNexis and Peekyou.

But this is just the tip of the iceberg; the not-for-profit consumer advocacy organisation, Privacy Rights Clearinghouse, currently (in 2019) lists a total of 172 companies on its website that are active in that market, of which 21 operate in the sensitive criminal record data tracking area.

WebFX, the digital marketing agency, makes the following claims on its website for the current size of the data brokerage industry:

- *Today, there are over 4,000 data brokering companies worldwide. Acxiom, one of the largest, has 23,000 servers collecting & analyzing consumer data, Data for 500 million consumers worldwide, and up to 1,500 data points per person – and that's just one company.*
- *In 2012, the data brokering industry generated $150 billion in revenue – that's twice the size of the entire intelligence budget of the United States government. Now, data brokering is a $200 billion industry, and it isn't showing any signs of becoming any less profitable.*

(WebFX, 2019)

There is also some evidence of consolidation occurring in the industry whereby some of the larger data brokerage businesses are owned by even larger corporations such as Reed Elsevier and Infosys Technologies. For example, in 2008 Reed Elsevier which already owned one of the major players, LexisNexis, also acquired the ChoicePoint data brokerage firm (Roderick, 2014).

The Regulatory Landscape

Despite their financial size and the level of influence their work can have on the lives of individuals and on society in general, the data brokerage industry is subject to little or no legal or regulatory oversight and it does all that it can to ensure that this remains the case. While it has been subjected to some scrutiny in the US – like the 2 inquiries that were undertaken by the US Congress in 2012 (Roderick, 2014) and in Canada by the Canadian Internet Policy and Public Interest Clinic (CIPPIC, 2016) and sponsored by the Office of the Privacy Commissioner of Canada - those exercises have resulted in only limited legal or regulatory change which is largely due to the uncooperative part that the data brokerage industry played in them. Justified or not, this is why the ethical 'alarm bells' have been ringing in relation to that industry for some time and will continue to do so.

In its struggle to avoid new regulatory controls the data brokerage industry has been able to take advantage of the disjointed nature of the relevant legal and regulatory frameworks that exist across, and even within, jurisdictions. In the United States, for instance, the federal legislation that could be used to underpin the regulation of data brokers is segmented numerous pieces including the Fair Credit Reporting Act (FCRA), Equal Credit Opportunity Act (ECOA), Fair Housing Act (FHA), Gramm-Leach-Bliley Act (GLBA) and the Dodd-Frank Act (DFA) – Rieke et al (2016), Roderick

(2014) – whose potential effectiveness is compromised further by the different implementation approaches taken by the different states. With the relatively recent enactment of the General Data Privacy Regulation (GDPR) the European Union is in a much stronger position to enact effective controls on the industry but even there differences in the national implementations of GDPR may frustrate efforts in this area too.

The only thing that the body of legislation in those two major jurisdictions have in common is their basis, to a greater or lesser extent, on the Fair Information Practice Principles (FIPPs) which cover rights and obligations in the following areas: notice, consent, choice, security, integrity, access and accountability (Kitchin, 2014). The FIPPs declare that individuals have the right to be informed when data is being collected about them and of the use to which is to be put and which point that person has to right to give or withhold his or her consent. They also put obligations on the acquirers of that data to ensure that it is accurate, is used only for the stated purpose, is retained for only the length of time needed to achieve that purpose and, while it is retained, that it is protected from unauthorised access and use.

However, it would not be fair to say that the regulatory landscape is uniformly bleak across the data brokerage industry in which the credit checking and scoring area is a relative bright spot and where Equifax, Experion and FICO are major players. The following factors may help to explain why this is the case:

1. It is one of the oldest types of data brokerage business areas.
2. It involves a lot of money (in the form of credit) and therefore is of particular fiscal interest to governments.
3. It is in the financial sector and therefore subject to that segment's more stringent regulatory environment – especially in the wake of the subprime mortgage-induced economic crash in 2008.

In fact, it is claimed that the credit rating data brokerage business itself was complicit in the creation of that crash by helping its customers to identify and target people who would be open to taking on risky mortgages at unsustainable subprime rates (Roderick, 2014).

In the EU, advances in the regulation in the credit rating area subsequent to the economic crash in 2008 have since permitted the development of transnational services in accordance with the enhanced regulatory controls (Rieke et al, 2016). The following diagram provides a snapshot of where that process was – and was aiming for – in 2010.

Figure 1: Source: Association of Consumer Credit Information Suppliers, ACCIS survey 2010 (Rieke et al, 2016)

The Big Picture

Before we look at what the ethical concerns hinted at earlier in this article might be, we need to get a better picture of what data brokers do, for and with whom and with what data from where. The following useful diagram provides a simplified but effective overview of the industry's key players and the data flows between them and shows the central part that the data brokers themselves play in it.

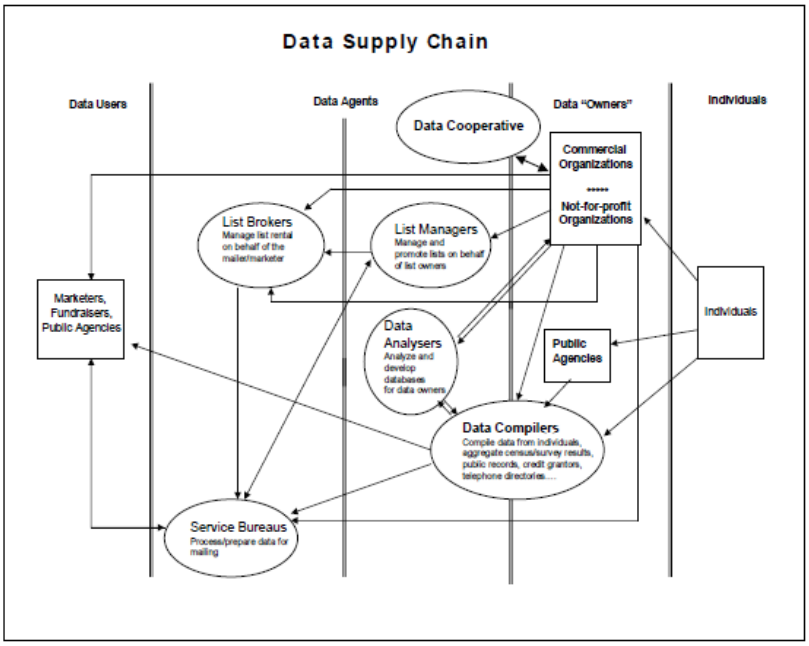

Figure 2: Key players and data flows (Source: The Canadian Internet Policy and Public Interest Clinic (2006))

That schematic shows how data captured from individual consumers (with or without their knowledge) in the Individuals swim lane on the far right is delivered to the Data Users swim lane on the extreme left. The data brokerage industry is represented by the two intervening Data Agents swim lanes where the Data Owners' data is aggregated (compiled), analysed and transformed, managed, marketed and sold (usually on a price-per-use rental basis). The diagram also shows how Data Owners sometimes bypass the Data Agents by selling their data directly to the Data Users and how this is sometimes done by groups (cooperatives) of Data Owners who pool their data before selling it, thereby enhancing its value.

To Regulate or Not to Regulate?

A report cited previously in this paper opines that '[p]otential uses of this data are limited only by law and ethics' (CIPPIC, 2014). The former approach means regulation and the latter one means self-regulation effectively. As discussed previously in this paper, the vast amounts of revenue that the industry is currently earning (and has been doing for some time), combined with human nature, means that the temptation to become 'economical with the ethics' must be very strong for some, if not many, data brokerage firms. Unfortunately for the majority of ethical operators, the very real risk that some of their peers will succumb to that temptation means that regulation is the only viable approach in the medium to long term.

Of course, the 'market solution' (Kitchin, 2016) promoted by some apologists of the data brokerage industry argue that 'market forces' is a third approach that could be taken and is the one that the industry itself favours. In other words, they contend that if a particular data brokerage does not manage its data and business properly the resultant data products will not be of sufficiently high quality to be successful in the marketplace and would therefore be supplanted by its competitors eventually.

However, one would have to ask the following questions whose answers are likely to effectively counter that argument:

1. How much damage would be done to individuals whose data is improperly or incorrectly processed in the meantime?
2. How would a particular Data User know that some, if not much, of the data supplied to them is of inferior quality if the results they achieve with the complete dataset (or list) are statistically better than those that they had been achieving without it?

Areas of Ethical Concern

As unethical behaviour by one person or organisation inevitably results in the unfair treatment of another one, this study will now use the Fair Information Practice Principles (FIPPs) to discuss the ethical issues that the continued lack of regulation of the data brokerage industry is like to spawn.

Notice

To make an informed decision in relation to consent (see below) an individual must firstly be made aware of what data are to be collected by the Data Owner and for what purpose.

In the context of the data brokerage industry, this type of information is usually provided in either the Terms of Service or Privacy Policy sections of the Data Owner's website and is often provided using either vague or arcane language. In many cases a clause is included that enables the Data Owner to change those provisions in the future without needing to inform the individual or to seek his or her consent.

Choice

The individual should have the right to opt in or opt out of the arrangement as notified to him or her under the provisions of the previous principle.

However, at the beginning of the transaction involving data brokerage context this is not a real choice at all because opting out invariably means that the provision of the relevant product or service will be refused. Thereafter, the Data Owner and/or the Data Agent typically make the process of opting out, if indeed such a facility is offered at all, so difficult that the most individuals who wish to avail of it eventually do not (or cannot practically) avail of it. In some case, where opt-out is supported, the individual may be forced to disclose additional person information to prove their identity which can be of use to the data broker before the opt-out process is completed (Roderick, 2014).

Consent

The individual must provide consent to the capture of the data relating him or her as specified during notification as well as to the sharing of that data with specified third parties and for what purpose.

As in the case of the Notice principle, this information is typically provided in the same arcane or vague terms in the same sources. Also, sharing of data will inevitably make ensuring its security and integrity more difficult which, in turn, increase the risk that it will be compromised.

Security

Protecting information from loss or misuse can be difficult to achieve especially if/when it is shared with one or more third parties.

In this context, the data brokerage industry has proven to be as susceptible to data breaches as any other sector. For instance, in 2005 the ChoicePoint brokerage revealed a data breach that involved 'the personal data of more than 163,000 US consumers' (Roderick, 2014). The US Federal Trade Commission (FTC) forced the company to pay out a total of $15m while making a profit of $148m from revenues of $1b that same year.

As such unauthorised data access can lead to identity theft which can have devastating consequences on the lives of the impacted people, in the case of the example cited above regulation failed to ensure that the punishment fitted the (potential) crime.

Integrity

The failure of a Data Owner or Data Agent to ensure the accuracy, currency and completeness of the data that it holds for, or derives about an individual can have extremely serious consequences for an individual ranging from wrongful arrest to disenfranchisement, or refusal of credit, or inability to gain employment etc. Yet the quality level of data held by data brokerages is being constantly called into question as typified by the following extract from an article on the Forbes website:

- *As with Oracle's data, Acxiom's profile on me was laughably inaccurate (83 out of 113 fields, 73%, were wrong) and in many cases the sources of information made no sense. According to the company I am married (false, but it said it got this from public records, retail activity and self-reporting) and I have three adults living in my apartment (again, false, but it claims I self-reported this at some point). I apparently purchased a $750,000 home in 2015 (I wish), have all sorts of credit cards and credit lines that were news to me, subscribe to services I didn't even know existed, love wine (I don't drink wine) and even own a dog I had no idea existed. It's truly amazing how much you can learn about yourself that you never knew!* (Forbes, 2018)

But, in the case of the credit rating data brokers, even where the data is accurate, it could be argued that their method penalise too harshly individuals who have fallen on hard times temporarily (Rieke et al, 2016).

Furthermore, the increasing use of data analytics by data brokerages to derive (infer) information about individuals based on known details of that individual or other 'similar' individuals, using non-transparent ('black box') algorithms, is only going to exacerbate this issue and, therefore, gives rise to the same type of ethical concerns. Increasingly, using predictive analytics, data brokerages make judgements on what an individual is likely to do (or not to do) in the future rather than on what he or she has done in the past and this also gives rise to ethical concerns.

Access

Despite the credit checking example cited above, most data brokerages make it extremely difficult, if not impossible, to access that data with a view to checking its accuracy and having it corrected. Indeed, because of the secrecy and size of the brokerage industry, it is very hard to determine which of them actually holds data on an individual in the first place. Again, the damage that inaccurate data can do to a person calls into question the ethical basis for this type of behaviour.

Accountability

Although this FIPP places an obligation on a Data Owner to ensure that it complies with all of the other FIPPs, without a regulatory regime with regular auditing and onerous penalties for non-compliance there can be no accountability.

A Way Forward

In their paper on the metaphors of big data, Puschmann and Burgess (2014) point out that 'the role of data as a valued commodity is effectively inscribed (e.g., "the new oil"; Rotella, 2012), most often by suggesting physicality, immutability, context independence, and intrinsic worth'. This means that the extremely valuable data flows that the data brokerage industry gives rise to could be also be fiscally valuable to the economies of the countries whose citizens and governments provide that data in the first place. The value of those data flows has already been quantified at $200b per annum and rising in a previous section of this document.

Bringing those data flows within the tax net could provide the powers that be with the incentive needed to establish and maintain the type of regulatory oversight needed to measure those flows to ensure that the revenue owed was collected. The required quantitative measurements could also be extended to include some qualitative ones. Once one government (or group of governments) acts, it would be difficult for others not to follow suit. In turn, the level of scrutiny that this would entail could drive improvement in the overall performance of data brokerage industry and assuage the ethical concerns that surround it at present.

Of course, establishing the required infrastructure (financial, technical and human) would not be without cost to the relevant governments but that investment could be recouped from the resultant new revenue stream over time. While quantitative measurement of the data flows by governments should not pose privacy concerns to their citizens, care would need to be taken to ensure that the types of qualitative metrics needed could not give rise to such concerns.

Conclusion

Until governments across the world make the concerted efforts needed to fully implement and improve the data privacy regulations needed to control the operations of the data brokerage industry effectively, the unethical practices that this document has highlighted, and their potential for inflicting serious damage on the lives of individuals, will continue. However, inertia is this regard is likely to persist until and unless those regimes are sufficiently incentivised to act.

This paper argues that the ongoing fiscal benefits that taxing the industry's data flows that have been identified in a previous section could provide the impetus to putting the necessary regulatory infrastructure in place.

References

Forbes, Leetaru K. (2018) *The Data Brokers So Powerful Even Facebook Bought Their Data - But They Got Me Wildly Wrong* [online]. Available at: https://www.forbes.com/sites/kalevleetaru/2018/04/05/the-data-brokers-so-powerful-even-facebook-bought-their-data-but-they-got-me-wildly-wrong/#1f7d4d813107 (accessed 6 May 2019),

Kitchin R. (2014) *The Data Revolution.* London, UK: SAGE Publications Ltd.

Kitchin, R. (2016) *Getting smarter about smart cities: Improving data privacy and data security*. Data Protection Unit, Department of the Taoiseach, Dublin, Ireland.

Laney D. (2001) 3D Data Management: Controlling Data Volume, Velocity, and Variety. META Group. Available at: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (accessed 6 May 2019).

Privacy Rights Clearing House (2019) *Online Information Brokers and Your Privacy* [online]. Available at: https://www.privacyrights.org/blog/online-information-brokers-and-your-privacy (accessed 6 May 2019).

Puschmann C. and Burgess J. (2014) Metaphors of Big Data. *International Journal of Communication* 8: 1690–1709. Available at: https://ijoc.org/index.php/ijoc/article/view/2169/1162/ (accessed 6 May 2019).

Rieke, A., Yu, H., Robinson, D. and von Hoboken, J. (2016) *Data Brokers in an Open Society*. Washington, D.C.: Open Society Foundations. Available at: https://www.opensocietyfoundations.org/sites/default/files/data-brokers-in-an-open-society-20161121.pdf (accessed 6 May 2019).

Roderick, L. (2014) *Discipline and power in the digital age: The Case of the US Consumer data broker industry*, Critical Sociology 40(5): 729–746.

WebFX (2019) *What Are Data Brokers – And What Is Your Data Worth?* [online]. Available at: https://www.webfx.com/blog/general/what-are-data-brokers-and-what-is-your-data-worth-infographic/ (accessed 6 May 2019)