Data Brokers and Human Rights

Big Data, Big Business

Occasional Paper Series
Paper Number 6

November 2016

Data Brokers and Human Rights

Big Data, Big Business

November 2016

Institute for Human Rights and Business
34b York Way
London, N1 9AB, UK
Phone: (+44) 203-411-433
Email: info@ihrb.org
Web: www.ihrb.org

Cite as: Institute for Human Rights and Business (IHRB), "Data Brokers and Human Rights: Big Data, Big Business" (November 2016), available at: https://www.ihrb.org/focus-areas/information-communication-technology/data-brokers-big-data-big-business.

Acknowledgements: This Paper was written by Jennifer L. Bauer, with input from Lucy Purdon, IHRB ICT Project Manager and Haley St. Dennis, IHRB Communications Manager.

Image: Nasa Earth Observatory, "Earth at night", Flickr/nasamarshall/13230982754.

Contents

Executive Summary 4

1. Big Data and the Every-Day Consumer 5

2. The Big Data Analytics Model 7

3. Benefits and Risks Associated with Big Data Analytics 9

Consumer Benefits of Big Data 9

Risks of Big Data Analytics 10

4. The Role of Data Brokers 11

5. Consumer Awareness 13

6. Corporate Due Diligence & Remedy 15

Questions for Human Rights Due Diligence 15

Access to Remedy 19

7. Recommendations 20

Increase Transparency 20

Expand User Knowledge 20

Increase User Control 21

Regulate Data Use Rather than Data Collection 21

Regulate Data Brokers 22

8. Conclusion 24

Executive Summary

This is the sixth in a series of occasional papers by the Institute for Human Rights and Business (IHRB). Papers in this series provide independent analysis and policy recommendations concerning timely subjects on the business and human rights agenda.

This paper was written to explore the impacts of big data on fundamental rights, highlight current expert opinion, and provoke deeper discussion and critical assessment of the roles and responsibilities of companies in the collection, storage, processing and sharing of personal data.[1] Although this paper focuses on one aspect of the big data landscape – the use of individual personal data for commercial purposes – it is a useful point from which to unpack the complexities of big data, including the opportunities to help realise human rights and the challenges data controllers face regarding ownership, consent, transparency, accountability and trust.

The resulting picture is that of low public awareness; lack of transparency in the collection, storage, processing and sharing of personal data; and a lack of clarity over accountability. Moreover, there is little guidance available for companies seeking to navigate this issue. A small number of high profile companies are actively engaged in identifying the human rights implications of big data and associated business responsibilities. But there are some businesses, like data brokers, that benefit from the lack of transparency. These companies “mine” and sell data, but little else is known about how they operate. They play a huge role in the big data ecosystem, yet they are not household names. This paper sheds light on the role of data brokers in the big data landscape.

Big data analytics can provide numerous benefits to business and society, but questions remain over how to mitigate the accompanying risks to human rights and establish effective safeguards on the collection, storage, use and sharing of personal data. Many of the benefits, risks, challenges, and recommendations associated with big data analytics have been documented by privacy scholars.[2] This paper focuses on the impacts to economic and social rights, such as discrimination in credit, employment, insurance, and mortgage decisions, and how these can disproportionately affect vulnerable populations like minorities and the poor.

This paper provides a user-friendly overview of big data analytics, including how it is used and operated in business settings, before discussing the benefits and risks associated with big data. It then proposes several recommendations in line with the UN Guiding Principles on Business and Human Rights[3] through which governments and businesses can improve transparency and accountability.

[1] In June 2016, IHRB co-hosted a Wilton Park event, Safeguarding Rights in the Big Data Revolution, which brought together 45 experts from international experts from government, companies, civil society and academia to discuss the challenges of safeguarding rights such as privacy and non-discrimination in the use of big data. The aim of the conference was to further explore the opportunities and risks of big data in the ICT sector and beyond with regards to respecting rights, and to further understand the practical steps companies can take to ensure rights are respected. See further: https://www.ihrb.org/focus-areas/information-communication-technology/safeguarding-rights-in-the-big-data-revolution

[2] See generally, e.g., Mayer- Victor Mayer-Schönberger & Kenneth Cukier, “Big Data” (2014) (describing big data analytics, how they are conducted, and their benefits and dangers); Bruce Schneier, “Data & Goliath: The Hidden Battles to Collect Your Data and Control Your World” (2015) (describing how big data analytics are being used to control an individual’s environment); Nathan Newman, “How Big Data Enables Economic Harm to Consumers, Especially to Low-Income and other Vulnerable Sectors of the Population”, 18 Journal of Internet Law 11 (Dec. 2014) (describing how big data analytics harm poor and minority consumers).

[3] See generally Office of the United Nations High Commissioner for Human Rights, “Guiding Principles on Business and Human Rights: Implementing the United Nations ‘Protect, Respect and Remedy’ Framework”, Hum. Rts. Council, U.N. Doc. A/HRC/17/31 (2011), available at: http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf [hereinafter UN Guiding Principles].

1. Big Data and the Every-Day Consumer

You leave a digital data trail on a daily basis. You use a credit or debit card to fill your gas tank first thing in the morning. Then you commute to work, using an electronic toll collection system. Perhaps you buy a coffee on your way into the office using a rewards account. You use a newspaper mobile app to catch up on the news while you wait in line. Later, you grab a drink with co-workers during happy hour and then stop for groceries on the way home. You use a credit or debit card to pay for the drinks, but use cash and your loyalty card to pay for the groceries. You spend an hour during your commute home catching up with friends on a hands-free cell phone, and then log into various social media accounts to check your correspondence. Before you go to bed, you check your step tracker and see you have managed to surpass your goal of 10,000 steps for the day. Smiling, you settle into bed, using your phone to turn down your thermostat and turn out your lights,[4] satisfied with your successful, productive day.

You may not realise it, but there are numerous vendors that are also satisfied with your productive day. At least twelve vendors now have direct information about your life, including your credit card provider, toll collector, coffee company, media outlet, grocery store, cell phone service provider, social media accounts, step tracker, thermostat service, and your mobile home automation app. Each of these vendors tracks and collects data about your use of their services. This may seem innocuous at first. Who cares if a coffee company knows that you buy a coffee daily, or that the grocery store knows that you tend to buy more near the end of the month? It is not like you have anything to hide.[5] And they just use that information to distribute coupons anyway, right?

But that is not all that is done with the data, and what may first appear innocuous at an individual transaction-level may not be innocuous when such data is gathered and aggregated on a large scale during big data analytics. Suddenly, these vendors have months of transactions through which they can see behavioral trends over time. And they have a pretty accurate picture of you. They know that you commute to work every day via a toll road;[6] that you tend to buy regular – not supreme – gas at a local pump two or three times a week; that you buy coffee every day and read the news on the go; that you tend to buy gin and tonics when you go out to drink; that you have a habit of calling friends while on the road;[7] that you care about your

[4] Matthew Bolton, "How to Control Your Home with Your iPhone or iPad", *TechRadar* (May 23, 2013), available at: http://www.techradar.com/us/news/mobile-computing/tablets/how-to-control-your-home-with-your-iphone-or-ipad-1151155.

[5] This is a common retort to privacy concerns, particularly in response to national security efforts. For a full explanation of why this is a false debate, *see generally* Daniel J. Solove, "Nothing To Hide: The False Trade-Off Between Privacy And Security" (2011), available at:
http://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2092&context=faculty_publications

[6] Schneier, *supra* note 2, at 15 (EZPass' automatic payment system links our license plate number and credit card to our identity).

[7] This is possible because every cell phone user makes an implicit bargain that their service carrier can track their geographical location so he may make and receive phone calls. Schneier, *supra* note 2, at 1. As Bruce Schneier opined, "*This is a very intimate form of surveillance. Your cell phone tracks where you live and where you work. It tracks where you like to spend your weekends and evenings. It tracks how often you go to church (and which church), how much time you spend in a bar, and whether you speed when you drive. It tracks—since it knows about all the other phones in your area—whom you spend your days with, whom you meet for lunch, and whom you sleep with.*" *Id.* at 1-2. This information is so valuable that phone carriers often sell it to data brokers, who then "*resell it to anyone willing to pay for it.*" *Id.* at 2.

health because you track your steps; and that you tend to lower the thermostat at night[8] and go to bed at a consistent hour.[9]

These innocuous insights into your life are surprisingly valuable.[10] It is for this reason that most vendors retain the right – via their privacy policies and various user agreements – to sell that information to third parties. Once that data can be sold or shared, it can be aggregated and analysed to create a comprehensive snapshot of you as a consumer.[11] Data brokers, for instance, collect consumer information to better categorise consumers and sell consumer profiles to other companies for marketing purposes. Where this data may once have been collected for a primary purpose, such as feedback for product improvement, it can now be recycled for an unlimited number of "secondary uses". The same medical information an individual reveals to a health organisation for diet advice can later be used to assess one's health insurance premiums; debt information can be used by banks during mortgage applications or sold to pay day lenders and debt consolidation companies.[12] The sheer number of vendors involved makes it extremely difficult, if not impossible, to find out which companies sold or bought information about you.

Although the concept of data analytics and data brokers is not new, this practice has reached an unprecedented sophistication and scale with the introduction of big data; and the volume of personal data available and the reliance on computer algorithms[13] to make decisions has introduced new vulnerabilities susceptible to abuse.[14]

The problem is that most people are not familiar with the volume of personal data that is being gathered, how it is being collected and from what sources, the extent to which that data is being sold or shared, and how it could negatively affect them.

[8] *Id.* at 15 ("*The smart thermostat [like Nest] adapts to your behavior patterns and responds to what's happening on the power grid. But . . . it records more than your energy usage: it also tracks and records your home's temperature, humidity, ambient light, and any nearby movement.*").

[9] *Id.* at 16 ("*If you wear a fitness tracking device like Fitbit or Jawbone, it collects information about your movements awake and asleep, and uses that to analyze both your exercise and sleep habits. It can [even] determine when you're having sex.*").

[10] According to a 2012 study, the annual revenue of just nine market players in the data broker business (discussed in Figure 2) was approximately $426 million. Federal Trade Commission, "Data Brokers: A Call for Transparency and Accountability", (May 2014), pg 23 [hereinafter FTC Data Broker Report]. But the data broker business actually has thousands of players. Steve Kroft, "The Data Brokers: Selling Your Personal Information", CBS News (Mar. 9, 2014), available at: http://www.cbsnews.com/news/the-data-brokers-selling-your-personal-information/. Other sources say that the global surveillance industry, which constitutes only a subset of how big data analytics can be used, produces an estimated revenue of $5 billion a year. Business & Human Rights Resource Centre, "Information Technology: The Power and Responsibility of Business" 3 (Feb. 2014) [hereinafter *The Power and Responsibility of Business*], available at: https://business-humanrights.org/sites/default/files/media/documents/information-technology-briefing-feb-2014.pdf.

[11] *Id.* at 20 ("*Governments and corporations gather, store, and analyze the tremendous amount of data we chuff out as we move through our digitized lives. Often this is without our knowledge, and typically without our consent. Based on this data, they draw conclusions about us that we might disagree with or object to, and that can impact our lives in profound ways.").*

[12] Melanie Hicken, "Big Data knows you're broke", *CNN Money* (29 April 2014), available at: http://money.cnn.com/2014/04/29/pf/data-brokers/index.html.

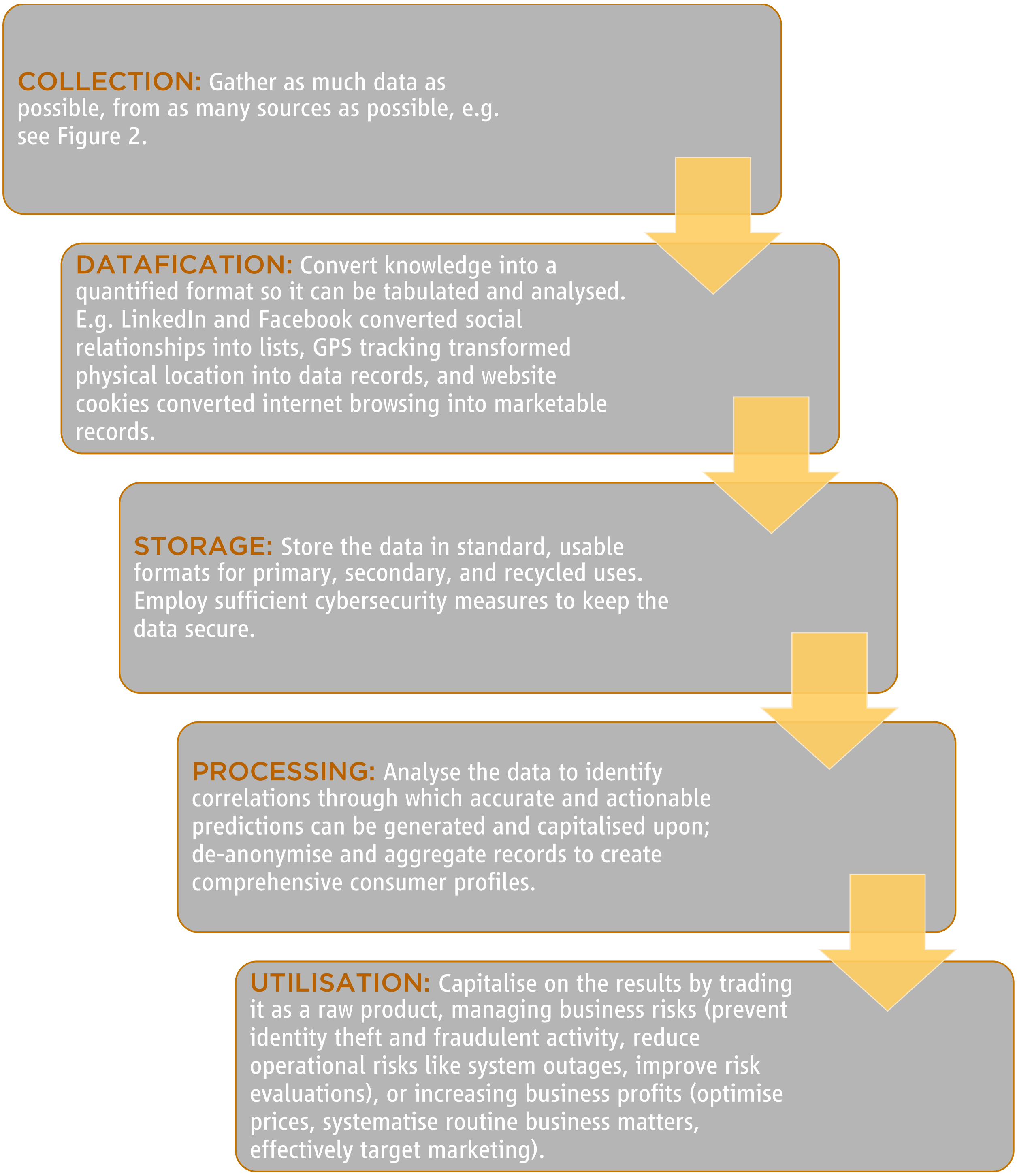
[13] An algorithm is a set of instructions in computer code that leads to a problem being solved or a decision being made. Algorithms are an essential part of big data analytics, which make decisions instead of a human, often about humans. It is unclear how the majority of algorithms work, and there is often a lack of transparency in how decisions are made, and therefore how decisions can be challenged.

[14] In 2014, the Federal Trade Commission (FTC) published a report warning about the practices of data brokers and called for more protection for transparency and accountability. *See generally* FTC Data Broker Report, supra note 10.

2. The Big Data Analytics Model

Big data analytics is a tool through which businesses analyse large, complex data sets to identify correlations, produce actionable predictions, and capitalise on the results by trading them as a raw product or using them to manage business risks, reduce operational risks, or increase profits. It is comprised of five steps – collection, datafication, storage, processing, and utilisation – outlined in in Figure 1.

Figure 1: The Big Data Analytics Model



COLLECTION: Gather as much data as possible, from as many sources as possible, e.g. see Figure 2.

DATAFICATION: Convert knowledge into a quantified format so it can be tabulated and analysed. E.g. LinkedIn and Facebook converted social relationships into lists, GPS tracking transformed physical location into data records, and website cookies converted internet browsing into marketable records.

STORAGE: Store the data in standard, usable formats for primary, secondary, and recycled uses. Employ sufficient cybersecurity measures to keep the data secure.

PROCESSING: Analyse the data to identify correlations through which accurate and actionable predictions can be generated and capitalised upon; de-anonymise and aggregate records to create comprehensive consumer profiles.

UTILISATION: Capitalise on the results by trading it as a raw product, managing business risks (prevent identity theft and fraudulent activity, reduce operational risks like system outages, improve risk evaluations), or increasing business profits (optimise prices, systematise routine business matters, effectively target marketing).

At first glance, this big data analytics model may seem unassuming. Businesses have collected and analysed consumer data to reduce costs and increase profits for years.[15] Data analytics is a common and accepted practice. But this established practice has reached an unprecedented sophistication and scale with the introduction of big data.[16] To understand this, it is helpful to view the variety of data incorporated in the data broker business model in Figure 2.

Big data differs from traditional data sets because of its **variety, volume,** and **velocity**.[17] Where traditional data sets were typically one-dimensional, big data sets incorporate everything from commercial data to government records and public data.[18] Commercial data is the component most analogous to traditional data sets because it includes all the data a business may hold on a customer. Amazon, for example, retains records on how often a customer uses its website (use-pattern information), how much they paid for a particular book (transaction data), their PayPal or credit card information (payment information), where these items were shipped (location data), and whether they decided to "follow" particular authors or add particular items to their "wish list" (website profile data). This is all considered commercial data. Where traditional data analytics focused perhaps solely on this type of data, big data analytics incorporates every data type and source. This includes government records, such as professional licenses recorded with the state, property and lien records, criminal records, census data, and voter registration information, and even public data like social media.[19] Because this data is dynamic and continuously produced from a variety of sources, it generates a high volume of data points at a fast rate or velocity.

It is these defining characteristics – variety, volume, and velocity – that allowed big data analytics to reach an unprecedented sophistication and scale.

[15] Kroft, *supra* note 12 ("Much of this is the kind of harmless consumer marketing that's been going on for decades.").

[16] *See id.*

[17] See Trevir Nath, "How Big Data has Changed Finance", Investopedia (last visited April 25, 2016), available at: http://www.investopedia.com/articles/active-trading/040915/how-big-data-has-changed-finance.asp.

[18] *See infra* Figure 2.

[19] FTC Data Broker Report, *supra* note 12, at 11-12.

3. Benefits and Risks Associated with Big Data Analytics

Consumer Benefits of Big Data

There are numerous benefits that big data analytics offer businesses and society overall. Big data analytics can help businesses increase profits by optimising prices,[20] improving the accuracy of targeted marketing,[21] and systematising routine business matters like performance reviews and employee management.[22] It can also reduce risk by preventing identity theft and fraudulent activity,[23] improving credit and insurance risk evaluations,[24] and preventing system outages.[25] Big data analytics also offers many societal benefits, such as the ability to create convenient and innovative tools that can track flu trends in real time, conduct medical research, identify and manage local fire hazards, and even help retirees live in their homes longer.[26] These innovations are likely a result, at least in part, of the fact that big data analytics create an efficient market in which market players share existing data rather than expending research and development funds to personally collect it.

[20] Newman, *supra* note 2, at 11-12 (explaining how big data analytics permits businesses to optimise prices by gathering data on consumer demographics and historical purchases to determine the maximum price those individuals are willing to pay).

[21] Veronica K. McGregor et al., "Big Data and Consumer Financial Information", *Business Law Today* (Nov. 2013), available at: http://www.americanbar.org/publications/blt/2013/11/04_mcgregor.html (explaining how the ability to collect consumers' purchase history through big data analytics means that companies can "track customer purchases back to the originating source" and identify the marketing strategies that have been successful). Businesses can also contract sophisticated target marketing tools through external data brokers; Experian, for example, offers a "Prospect Triggers" service that "targets consumers nationwide based on their actual credit behavior within the past 24-48 hours." Ed Mierzwinski & Jeff Chester, "Selling Consumers Not Lists: The New World of Digital Decision-Making and the Role of the Fair Credit Reporting Act", *46 Suffolk U. L. Rev.* 845, 855-56 (2013), available at: http://suffolklawreview.org/wp-content/uploads/2014/01/Mierzwinksi-Chester_Lead.pdf.

[22] Jennifer Alsever, "Is Software Better at Managing People than You Are?" *Fortune* (21 Mar. 2016), available at: http://fortune.com/2016/03/21/software-algorithms-hiring/. A Cisco-design management tool called Team Space, for instance, uses a proprietary algorithm to analyse employee surveys to determine each employee's strengths, how he works best, and what energises him so the manager can receive tailored management advice. *Id.* Another program, WideAngle, offers support for improving one-on-one management meetings. *See, e.g.*, Wide Angle, https://wideangle.com (last visited April 24, 2016). IBM uses a data analytics program called Blue Matching to match employees to new opportunities internally based on their role, skill, experience, location preference, and performance. Alsever, *supra* note 21. Companies are investing significant funds into these predictive HR analytical tools, investing $2 billion in companies producing these apps during the past year alone, and several big companies like General Electric, Accenture, Deloitte, Adobe, Dell, IBM, and Microsoft have broadly adopted them. *Id.* Although employee management and HR tasks are routine business matters that all businesses share, there are many other routine business matters that may benefit from being systematised even though they vary across industries. Algorithmic trading, for instance, is simply proprietary algorithms designed to systematise financial trading via big data analytics. Xiongpai Qin, "Making Use of the Big Data: Next Generation of Algorithm Trading", *SpringerLink*, http://link.springer.com/chapter/10.1007/978-3-642-33478-8_5?no-access=true (last visited April 24, 2016) ("*Algorithm trading is using computer programs to automate trading actions without much human intervention. . . . The soul of algorithm trading is the trading strategies, which are built upon technical analysis rules, statistical methods, and machine learning techniques.*"). Most large financial institutions likely build such algorithmic trading models internally, but there are several vendors prepared to troubleshoot any errors or difficulties such large businesses (or even some smaller businesses interested in gaining a competitive edge) encounter. *See, e.g.*, DataArt, (last visited April 24, 2016).

[23] Big data analytics enable financial institutions to reduce the risk and costs associated with identity theft and fraudulent financial activity by identifying when transactions deviate from established spending patterns. *See* McGregor, *supra* note 20. Institutions use big data analytics to create a customer profile based on "customer characteristics and [historical] spending patterns," and then compare those to all future activity. *Id.* The financial institution can then issue a fraud alert once it notices suspicious activity, and freeze the account to prevent further fraudulent activity and reduce overall costs.

[24] *See, e.g.,* Data Scoring, (last visited May 27, 2016).

[25] The financial loss and reputational damage that could result from a trading platform outage are significant. Goldman Sachs may have lost as much as $100 million during a recent 15-minute outage of its options desk; Knight Capital lost more than $400 million when it lost a single server for 45 minutes. David Lauer, "The Challenges and Opportunities of Big Data in Finance", *Inside Market Data* (2 Dec. 2013), available at: http://www.waterstechnology.com/inside-market-data/opinion/2310216/open-platform-the-challenges-and-opportunities-of-big-data-in-finance (subscription needed). According to a SEC review of the Knight Capital incident, the financial institution should have been aware of the software problems because the institution had experienced a similar trading issue in the past and this particular incident was preceded by a flood of emails. *Id.* Big data analytics permit businesses to monitor for such indicators and use case-based reasoning to prevent outages or minimise downtime. *Id.*

[26] Mayer-Schönberger & Cukier, *supra* note 2, at 1-2 (2014) (Google can track the H1N1 virus in live time); Craig Mundie, "Privacy Pragmatism", *Foreign Affairs*, (March/April 2014), available at: https://www.foreignaffairs.com/articles/2014-02-12/privacy-pragmatism ("*[I]n 2011, researchers at the health-care giant Kaiser Permanente used the medical records of 3.2 million individuals to find a link between autism spectrum disorders in children and their mothers' use of antidepressant drugs.*"); Mayer-Schönberger & Kenneth Cukier, *supra* note 2 (detailing a project in which local officials identified fire hazards using big data analytics and prioritised repairs accordingly); The Week Magazine, "Technology *for Retirees:* Innovations that will Help Retirees Live Longer, Healthier, More Active Lives", (22 April 2016), pg. 22 (reporting that the "*Internet of Things will help retirees stay in their homes longer*" by producing "*[w]eb-connected, smartphone-controlled home devices [that] will give seniors a hand in running the house: automatic lighting that turns on and off at programmed hours, smart doorbells that show who is knocking, and appliances that know to shut off when no one is in the house*").

Risks of Big Data Analytics

Big data analytics may permit businesses to optimise prices, but it also runs the risk of creating a market where people are exploitatively profiled. For example, offered higher prices whilst unaware that other customers are getting better deals – as a result, financially struggling houses are tagged as vulnerable and offered economically exploitative services such as payday and subprime loans.[27] There is already ample evidence that such attempts to maximise profits lead to price discrimination,[28] which has a negative effect on the entire economy because it increases the overall cost of products within economic models whenever prices are obscured[29] and limits one's access to certain goods and services.[30]

Other discriminatory effects may arise when data is misinterpreted: Efforts to systematise routine human resource matters, for example, could reinforce pre-existing internal issues, like hiring discrimination, that is embedded within the data.[31]

Finally, because big data analytics is based on correlations, similarly discriminatory effects can hide behind seemingly unrelated characteristics. Wells Fargo, for instance, committed a modern version of "redlining" when its online calculator – marketed as a tool for potential mortgage customers to search for a new home – used the prospective homeowners' current zip code to direct them to a neighborhood that reflected the dominant race of their current location.[32]

Other correlations, such as assessing one's creditworthiness based on the creditworthiness of those with which they associate,[33] may not yet be illegal but run the risk of producing impermissibly discriminatory effects. Big data analytics certainly offer businesses a means of increasing profits and reducing operational risks, but it is accompanied by diverse regulatory hurdles, increased cybersecurity costs, and liability for data breaches and discriminatory effects on consumers.

[27] Newman, *supra* note 2, at 11.

[28] This is the business practice of charging different prices to different people to maximise profits. Schneier, *supra* note 2, at 109. Discover Financial Services was one of several major companies discovered to have used user location data to systematically display different online prices to different customers in 2012. Newman, *supra* note 2, at 12. Capital One similarly displays different credit card deals depending on the viewer's location and estimated income. *Id.* Staples, Rosetta Stone, and Home Depot display different website prices depending on where the viewer is located and how close a competitor's store is located. Schneier, *supra* note 2, at 110. Orbitz highlighted different hotel room prices based on whether viewers were using Mac or Windows. *Id.* at 111.

[29] Newman, *supra* note 2, at 13. According to one study, the use of price discrimination to increase profits raised "prices overall, with many consumers paying twice as much as others for the same product." *Id.*

[30] Schneier, *supra* note 2, at 111.

[31] Alsever, *supra* note 21.

[32] Schneier, *supra* note 2, at 109. "Redlining," or the practice of "denying or charging more for services by using neighborhood as a proxy for race," is illegal. *Id.*; *see also* Federal Reserve, "Federal Fair Lending Regulations and Statutes Fair Housing Act", Consumer Compliance Handbook, (last visited April 24, 2016), available at: https://www.federalreserve.gov/boarddocs/supmanual/supervision_cch.htm.

[33] American Express, for example, has reduced credit limits because the individual shopped at stores where the majority of customers had poor credit.[33] Schneier, *supra* note 2, at 111. at 113. Another company, Lenddo, uses the creditworthiness of those one frequently interacts with on Facebook. *Id.* at 113. The practice was originally designed so banks could give loans to people without credit ratings, but it is now much more widespread. *Id.* As one expert explained, "[i]f your habits associate you with particular categories or groups, [despite your right to freely associate with anyone you like,] you will invisibly find opportunities opening up or closing down based on how data algorithms choose to place you." Newman, *supra* note 2, at 14.

4. The Role of Data Brokers

The sophistication and scale inherent in successful big data analytics has enabled data brokers to market it as a process that other businesses should outsource.[34] The data broker collects, aggregates, and de-anonymises all the data identified in Figure 2 below, and then sells the resulting consumer profiles as a raw product (which the receiving business can then incorporate and use as they see fit) or a service (to companies that do not wish to spend time determining how to use the data or updating it periodically). Facebook, for instance, contracted with data brokers Acxiom and Epsilon to match online user profiles with in-store purchases.[35] Starbucks, Kmart, and Subway similarly contracted with Placecast to use geo-fencing – a new big data technique where marketers identify people near a particular businesses – to deliver ads to the cellphones of nearby customers (reaching approximately ten million phones in the US and the UK).[36]

The benefit of outsourcing this analytical process to a data broker is that data brokers can do the legwork of buying data from a variety of commercial sources and combining it with information collected from government records and other publicly available sources. The commercial data a data broker can compile is particularly expansive, including records from retailers, catalog companies, registration websites (like news or travel websites), telephone companies, automobile dealers, and even some financial services.[37] It even encompasses brick-and-mortar store transactions that are translated into digital records.[38] The practice of individual retailers, both online and brick-and-mortar stores, selling their transaction data is so common that just *nine* data brokers in 2012 had amassed the purchase history of over 190 million consumers from more than 2600 merchants.[39]

However, data brokers offer the additional benefit of aggregating the data and, depending on how the contracting company requested the data, presenting it in a way that is immediately valuable. For example, data brokers often compile and sell descriptive customer lists like "dog owner", "winter activity enthusiast", "married sophisticate", "expectant parent", and "urban scramble" to companies for target marketing.[40]

[34] FTC Data Broker Report, *supra* note 12, at i (defining data brokers as "*companies that collect consumers' personal information and resell or share that information with others*"). Contracting parties include other companies, data brokers, and even the government.

[35] Schneier, *supra* note 2, at 41.

[36] *Id.* at 39.

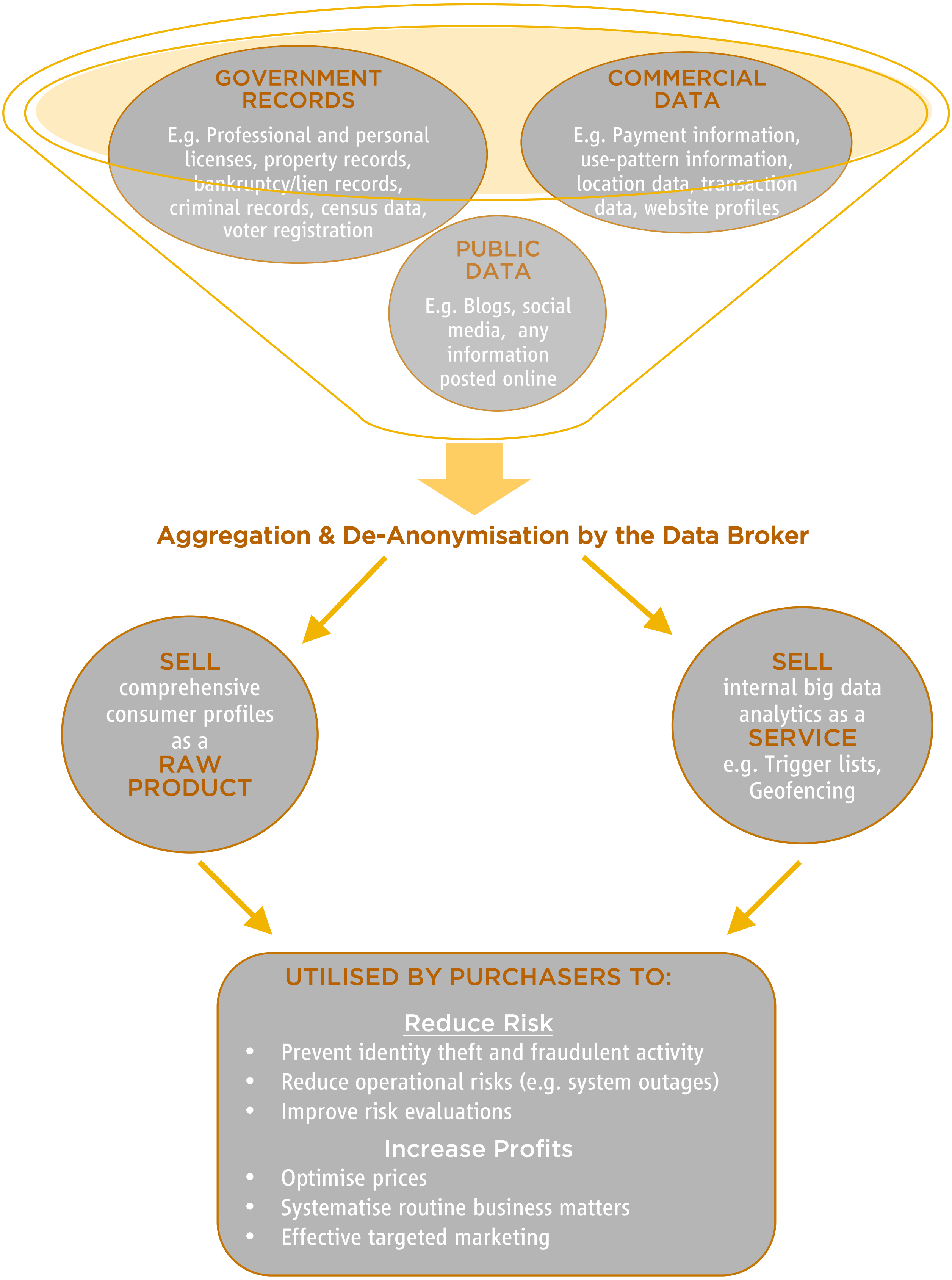
[37] FTC Data Broker Report, *supra* note 12, at 13-14.

[38] 60 Minutes Overtime Video, "Shocked to Learn How Data Brokers Are Watching You?", *CBS News* (9 Mar. 2014), available at: http://www.cbsnews.com/news/shocked-to-learn-how-data-brokers-are-watching-you/ (indicating that the recent explosion in this industry is because of the convergence of online and offline data).

[39] FTC Data Broker Report, *supra* note 12, at 14.

[40] *Id.* at 47. Still other data brokers identify and sell lists of consumers that are gay or lesbian, suffering from bipolar disorder, addicted to alcohol or gambling, chronically in debt, contracted a sexually transmitted disease, or even purchased adult material or sex toys. Kroft, *supra* note 12.

Figure 2: The Data Broker Business Model



GOVERNMENT RECORDS
E.g. Professional and personal licenses, property records, bankruptcy/lien records, criminal records, census data, voter registration

COMMERCIAL DATA
E.g. Payment information, use-pattern information, location data, transaction data, website profiles

PUBLIC DATA
E.g. Blogs, social media, any information posted online

Aggregation & De-Anonymisation by the Data Broker

SELL comprehensive consumer profiles as a RAW PRODUCT

SELL internal big data analytics as a SERVICE e.g. Trigger lists, Geofencing

UTILISED BY PURCHASERS TO:

Reduce Risk

- Prevent identity theft and fraudulent activity
- Reduce operational risks (e.g. system outages)
- Improve risk evaluations

Increase Profits

- Optimise prices
- Systematise routine business matters
- Effective targeted marketing

5. Consumer Awareness

Consumers are largely unaware of how businesses are buying, consolidating, aggregating, and analysing their data.[41] Realising their ability to generate revenue by selling their customer data to data brokers, some companies are designing their service platforms to collect as much data as possible.[42] Popular smartphone applications "Angry Birds" and "Brightest Flashlight Free," for example, collect location data even when they are closed.[43] Other applications download copies of the user's address book, calendar, bookmarks, and search history.[44] Credit rating agencies offer "trigger lists" where they monitor for credit report inquiries that signal an interest in a particular financial service (like refinancing one's mortgage) and sell that interest to a competing financial service provider.[45] Other companies operate online communities, like "GoodParentingToday.com" and OKCupid's dating website, to collect and sell information posted by users.[46] Many more use cookies to track user activity online.[47]

The common defense raised against any critiques of the system is that consumers are aware of the bargain they make in giving up some privacy in exchange for the benefits of using the internet.[48] The terms of that bargain are outlined in each website or company's privacy policy.[49] Professor Daniel J. Solove called this idea – that individuals can consent to the collection, use, or disclosure of their personal data – "privacy self-management."[50] Under the privacy self-management framework, consumers have legal

> *"rights to notice, access, and consent regarding the collection, use, and disclosure of personal data"* that they can choose to exercise after *"weigh[ing] the costs and benefits of the collection, use, or disclosure of their information."*[51]

[41] *Id.*; *see also* FTC Data Broker Report, *supra* note 12, at 46.

[42] Schneier, *supra* note 2, at 3. This differs from companies that first use the data to offer a particular service: Google Maps uses location data to provide directions, Uber uses it to identify pick-up locations and nearby drivers, and Yelp uses it to identify nearby restaurants and bars. *Id.* Even then, companies may be able to recycle the data for a secondary use, such as selling it to a data broker, depending on what their user agreement and privacy policy state. Kroft, *supra* note 12 ("Most retailers are finding out that they have a secondary source of income, which is that the data about their customers is probably just about as valuable, maybe even more so, than the actual product or service that they're selling to the individual."). Others, such as data broker "Take5Solutions," operate websites like "GoodParentingToday.com" and "T5 HealthyLiving.com" where users join online communities to share stories about their families and health and the underlying company collects and sells this information. *Id.*

[43] *Id.*; Schneier, *supra* note 2, at 3.

[44] *Id.*

[45] Experian, for example, offers a "Prospect Triggers" service that "*targets consumers nationwide based on their actual credit behavior within the past 24-48 hours.*" Mierzwinski & Chester, *supra* note 20, at 855-56.

[46] Kroft, *supra* note 12.

[47] Cookies are persistent identifiers that are installed when you first visit a website. Schneier, *supra* note 2, at 47. They were first used to recognise customers and personalise their website experience, perhaps by retaining one's shopping cart. *Id.* at 47-48. But they now track one's click activity on a website—including "*where they click, how long they look at the page, where the mouse-cursor hovers, what they type, and more*"—and can even track the user's activity across websites. *Id.* at 48; Mayer- Mayer-Schönberger & Cukier, *supra* note 2, at 113. Facebook and Google, for instance, can track users wherever a Facebook Like or Google Plus button exists, even if the user is not logged into either service. Schneier, *supra* note 2, at 48.

[48] 60 Minutes Overtime Video, "A Data Broker Defends His Industry", *CBS News* (9 Mar. 2014), available at: http://www.cbsnews.com/videos/a-data-broker-defends-his-industry/ (indicating that consumers vote with their feet and willingly share this information on the internet). *Contra* Newman, *supra* note 2, at 17 (citing a 2012 Pew survey as evidence that many consumers wish they had greater control over how their data is used: according to the survey, 73 percent of the US public oppose having their internet search history tracked and used to improve search engine results, and 68 percent oppose using user data for target advertisements).

[49] *See* Kroft, *supra* note 12. Visit the legal section of the OKCupid website, browse the terms and conditions, and within the privacy policy is a single line stating that "[y]ou should appreciate that all information submitted on the Website might potentially be publicly accessible." *Id.*

[50] Daniel J. Solove, "Privacy Self-Management and the Consent Dilemma", 126 *Harv. L. Rev.* 1879, 1880 (2013), available: http://paulohm.com/classes/infopriv13/files/week4/ExcerptSolovePrivacySelfManagementAndConsentDilemma.pdf.

[51] *Id.*

But privacy self-management is ineffective, and the notice and consent framework upon which it depends is flawed. As the President's Council of Advisors on Science and Technology (PCAST) summarised in its 2014 presidential report:

> *"[i]n some fantasy world, users actually read these notices, understand their legal implications (consulting their attorneys if necessary), negotiate with other providers of similar services to get better privacy treatment, and only then click to indicate their consent. Reality is different."*[52]

In reality, people do not read privacy policies.[53] Even if they did, they often do not know enough to understand them or make an informed choice.[54] Nor can the user exercise meaningful control in such a manner (even if one did read and understand every privacy policy) because the very benefit big data provides in creating *"new, non-obvious, unexpectedly powerful uses of data"* defeats notice and consent as an effective policy tool: *"[i]t is simply too complicated for the individual to make fine-grained choices for every new situation or app."*[55] Moreover, the risks of divulging such information cannot be *"adequately assessed in a series of isolated transactions"* because its aggregation by various data brokers has a cumulative effect that a user cannot reasonably anticipate or calculate at the time of collection.[56]

The scale and sophistication that big data analytics have reached with the rise of the data broker business model is so high, and the resulting network so complex, that it is virtually impossible for any consumer to identify how a data broker obtained his or her data.[57] Nor is there a consistent way for consumers to access, correct, or suppress it.[58] If privacy self-management were a truly feasible means of managing data use and control, then it would be fairly straight-forward for an individual to opt out of these practices. But it is far from easy. Indeed, a journalist once spent a year attempting to avoid online trackers, only to find that it was hopeless despite expending significant time and effort and using the best technological tools available.[59] She documented the experience in her book Dragnet Nation.[60] Her practical experience exemplifies how the three common strategies for managing privacy in the big data era – individual notice and consent, opting out, and anonymisation[61] – have *"lost much of their effectiveness."*[62]

[52] President's Council of Advisors on Science and Technology, "Report to the President – Big Data and Privacy: A Technological Perspective", (May 2014), pg 38, available at: https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf [hereinafter PCAST Presidential Report]. Consider, for example, that many business take advantage of a user's limited bargaining power by "*conditioning products, services, or access on opting in*" and making it common for "*agreeing to . . . end-user license agreements . . . [to be] a perquisite for obtaining access to a website or to use a product or service.*" Solove, *supra* note 30, at 1898. A consumer's limited bargaining power is particularly apparent when one considers that: Most privacy policies provide no way for customers to prevent changes in the policy, and they lack a binding enforcement mechanism. Frequently, companies revise their privacy policies, making it even more difficult for an individual to keep track…Further, personal information databases can be sold to other businesses with less protective privacy policies, especially when a company goes bankrupt. Daniel J. Solove, *The Digital Person: Technology and Privacy in the Information Age*, (2004), pg 83.

[53] Solove, *supra* note 50, at 1888.

[54] *Id.*

[55] PCAST Presidential Report, *supra* note 52, at 38.

[56] Solove, *supra* note 50, at 1893.

[57] *Id.*

[58] 60 Minutes Overtime Video, "How to Defend Your Privacy Online", *CBS News* (9 Mar. 2014), available at: http://www.cbsnews.com/news/how-to-defend-your-privacy-online/; *see also* FTC Data Broker Report, *supra* note 12, at 50 (recommending legislation creating "*a centralized mechanism, such as an Internet portal, where data brokers can identify themselves, describe their information collection and use practices, and provide links to access tools and opt outs*").

[59] *Id*

[60] *Id.*

[61] Anonymisation – or the claim that the information is safe and does not endanger an individual's privacy – is another common defense against critiques of big data analytics. Kroft, *supra* note 12. But it is very easy for data brokers, after collecting a large amount of information, to de-anonymise and link the information by identifying a common denominator like an email address or IP address. *Id.* Indeed, a graduate student proved in 2000 that "*87 percent of all Americans could be uniquely identified using only three bits of information: ZIP code, birthdate, and sex.*" Nate Anderson, "'Anonymized Data' Really Isn't – and Here's Why Not", *Ars Technica* (8 Sept. 2009), available at: http://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/. Other research has shown that "*95% of Americans can be identified by name from just four time/date/location points.*" Schneier, *supra* note 2, at 44.

[62] Newman, *supra* note 2, at 19.

6. Corporate Due Diligence & Remedy

Companies whose business model relies on, or is indeed built around, the insights extracted from personal data, cannot outsource their responsibility to respect human rights to data brokers, and must conduct due human rights diligence as outlined in the UN Guiding Principles on Business and Human Rights.[63]

The most immediate way in which businesses can implement the UN Guiding Principles on Business and Human Rights and have an impact on the system is by establishing robust human rights due diligence processes.[64] Because these due diligence processes must encompass every stage of business operations and business relationships with suppliers and customers,[65] many of the issues within the current big data realm would be addressed.

Whether companies are collecting and analysing data themselves or employing the services of a data broker, a strong due diligence process focused on addressing human rights impacts within a business using big data analytics should ask the questions and address the points outlined below.

Many of the questions raised during these stages overlap with each other. For instance, a business may not wish to collect the data at all if it decides the cybersecurity risks and storage costs are too high. That is why these questions should be embedded within each department's operational procedures, incorporated into internal reporting, and addressed whenever a new customer or business proposal is considered by management.

Questions for Human Rights Due Diligence

Collection

- If data is purchased from another entity, how can the company be confident that it was legally acquired? Do we need to verify its accuracy? What parameters have we put in place internally to ensure that any data we collect ourselves is accurate and legally acquired?

- If the data is restricted in some way (e.g. it cannot be used for consumer evaluations under the Fair Credit Reporting Act (FCRA) in the USA), then how do we tag the data so these

[63] UN Guiding Principles *supra* note 3, *see also id.* § II.A, ¶ 11, at 13 ("*Business enterprises should respect human rights. This means that they should avoid infringing on the human rights of others and should address adverse human rights impacts with which they are involved.*"); *Id.* § II.A, ¶ 11, at 13 (accompanying commentary) ("*It exists independently of States' abilities and/or willingness to fulfill their own human rights obligations, and does not diminish those obligations.*").

[64] This would address the business responsibilities laid out in UN Guiding Principles 15-21.

[65] UN Guiding Principles, *supra* note 90, § II.A, ¶ 13, at 14 ("*The responsibility to respect human rights requires that business enterprises: (a) Avoid causing or contributing to adverse human rights impacts through their own activities, and address such impacts when they occur; (b) Seek to prevent or mitigate adverse human rights impacts that are directly linked to their operations, products or services by their business relationships, even if they have not contributed to those impacts.*"); *Id.* § II.A, ¶ 13, at 15 (accompanying commentary) ("*Business enterprises may be involved with adverse human rights impacts either through their own activities or as a result of their business relationships with other parties.*"); *Id.* § II.B, ¶ 17, at 18 (accompanying commentary) ("*[B]usiness enterprises may be perceived as being "complicit" in the acts of another party where, for example, they are seen to benefit from an abuse committed by that party.*").

restrictions can be honored later on in our operations? See also the below questions on data utilisation.

- Are our privacy policy and user agreements easily accessible and understood by the public? How clearly do they identify the restrictions our business places on data use and what we consider appropriate uses? Do they offer our users meaningful control over their data during and after collection?

- How difficult is it for users to opt-out of data collection, or opt-out of only particular uses (e.g. limit secondary uses like selling or sharing data with others)? Is data collection a pre-requisite for using our service or product, or can an individual opt-out of data collection (or opt-out of secondary uses) and still use our product or service?

- Is this data accessible to the user after collection?

  - One of the questions data collectors must decide at the point of collection is whether users will be able to access the data after collection, in what form it will be available, and perhaps even the level of security required to store it. For example, the European Union (EU) requires that EU citizens can access, verify, and correct their data, even after they give it to a company.

Datafication

- Should this knowledge be converted into data points that can be sold/reused (i.e. via datafication)?

  - Businesses should ask themselves whether the information they are seeking to collect – such as the personal and professional relationships that are detailed via Facebook and LinkedIn – should be datafied (converted into data points that can be sold/reused). Those personal and professional relationships were once ephemeral (an intangible fact that required significant effort and surveillance to identify/determine) and are now data points that LinkedIn and Facebook can easily share/sell to others. The early pioneers in this area could perhaps be forgiven for not understanding all the potential ramifications of datafying these ephemeral points, but businesses should now ask themselves whether the knowledge they are seeking/acquiring actually should be datafied.

- The company should consider the costs and benefits involved in converting and handling data:

  - How does the company intend to use it? Is there a way the data could be misappropriated and abused?
  - How securely would it need to be stored? How expensive would this storage and security be?
  - Does it make our business a target for increased cybersecurity threats, regulatory oversight, or judicial orders (e.g. search warrants)?

Storage

- How securely does this data need to be stored? How much would it cost?

  – Cybersecurity is similar to insurance in the sense that it is most effective when narrowly tailored to the customer's needs. Certain industries (e.g. healthcare, social security/welfare, the legal profession) rely on particularly sensitive data that requires a higher degree of security from theft and/or other data breaches. Currently, those security needs are often based on how commercially valuable the data is (e.g. how valuable merger and acquisition data would be, or stealing someone's social security number) but perhaps this evaluation should also consider the social value associated with particular data (e.g. the negative impacts on social rights).

- Does it increase the cybersecurity threats to the business (e.g. hackers)?

  – Lawyers, for instance, often have access to social security numbers, health information, credit card numbers, financial records (like taxes, property records, etc), lawsuits, and mergers and acquisitions. This information is valuable to the internal parties and others (like hackers) that want to steal it, use it, or otherwise compromise it.

- What are the threats to the user or other data subject in the event of a breach? What would the financial, legal, and regulatory ramifications of such a breach be?

- What internal review processes does our business have in place to periodically review the above decisions, particularly as the security requirements and cybersecurity risks associated with the data may change with the innovation of new technologies and data uses?

- How frequently do we audit our stored data and evaluate the need to continue storing it?

Processing

- How will this data be aggregated and processed? Will this data be de-anonymised?

  – For example, will consumer data be identified by a random consistent identifier, aggregated with other (external) data, and then processed so the consistent identifier is de-anonymised / linked to a name? The two data sets only need to share a single data point (e.g. both data sets have the same IP address or email address) to de-anonymise the data. Data brokers do this multiple times (combining multiple data sets based on common identifiers/data points) to build a comprehensive profile.

- What is the risk to the user or data subject if the data is de-anonymised and aggregated with other data connected to that person? See the below questions on how the data could be used, misappropriated, or abused.

Utilisation

- How does our business intend to use this data? What parties and/or individuals may be affected by this use? Could it negatively affect their human rights, such as their privacy, right to freedom of expression and association?

- Could this data be misappropriated for another purpose? How else could this data be used and, possibly, abused? What human rights could be implicated by these other potential uses? What precautions need to be implemented to ensure this data cannot be misappropriated in a manner that adversely affects human rights?

- How do we ensure that our business partners, including our customers,[66] do not misappropriate this data or use it to adversely affect human rights?[67]

  - Possible precautions include:

    - Customer due diligence: Does the customer have a history of human rights abuses or data misappropriation?[68]

    - Project due diligence: How does the customer intend to use this project? Does that purpose adversely affect anyone's human rights?

    - Precautions against misuse: Include contractual right to refuse upgrades or cancel service in the event of misuse, itemise or otherwise describe appropriate uses or explicit restrictions within the contract; install a remote-controlled "kill-switch" that can either terminate the program or disable certain functions if such an event should occur.[69] It may seem odd to be discussing contractual provisions in a step about uses/results, but the challenge is that none of the other stages – collection, datafication, storage, or processing – address this stage of forming client relationships (e.g. negotiating a contract or other terms of service). In this case, one of the best ways to resolve the issues involved in the utilisation stage – e.g. whether the data can be misused or otherwise abused – is best addressed when first forming client relationships and establishing the terms of service/client contract.

    - How, and for how long, do we monitor such customers and projects? How do we know when the above precautions, such as a remote-controlled switch or termination of service, should occur?

[66] Businesses may sometimes seek to excuse indirect human rights impacts by claiming that they cannot be responsible for the actions of their customers. Human Rights First, "Excuses, Excuses: Surveillance Technology and Oppressive Regimes", (18 November 2011), available at: http://www.humanrightsfirst.org/blog/excuses-excuses-surveillance-technology-and-oppressive-regimes. But this is incorrect because, whether or not the sale itself is illegal, the Guiding Principles specifically state that businesses can be responsible for the actions of third parties connected to their operations, products or services. *Guiding Principles*, *supra* note 85, § II.A, ¶ 13, at 14 ("*The responsibility to respect human rights requires that business enterprises . . . [s]eek to prevent or mitigate adverse human rights impacts that are directly linked to their operations, products or services by their business relationships, even if they have not contributed to those impacts.*"); Roggensack & Walters, *supra* note 93 ("*Whether the sale of lawful intercept technology to repressive regimes is technically illegal or not, companies should at a minimum exercise due diligence before making such sales given the capacity of this technology to facilitate expression.*").

[67] Businesses may also sometimes claim that they cannot be responsible for the misuse of their services or products. *Id.* ("*We sell our products to private companies, not to governments, and we have the expectation that our product will only be used for commercial purposes., not repression by a government.*"). But the reality is that if a "*sale is made to a company in a country with a record of repression, it carries the risk that the technology might be coopted by security forces seeking to misuse it, and companies have a due diligence obligation to investigate and avoid that possibility.*" *Id.*

[68] Often, this query will be particularly relevant when considering a transaction with another company or nation-state (rather than a retail consumer). *Id.*

[69] *Id.* ("*It is imperative that technology providers make every conscious and concerted effort to advance the benefits and minimize the abusive uses of their products. Providers may consider development of a remote (company-controlled) kill switch or a built-in ability to disable certain functions in the event of misuse. Additionally, companies could include a contractual right to terminate service or refuse upgrades for violating parties.*").

Access to Remedy

Businesses should also ensure that there are mechanisms for addressing human rights impacts raised at each stage of the above process, including an effective and accessible non-judicial mechanism for affected individuals to raise grievances.

Possible grievances that could be raised at every stages of the big data analytics model include:

- illegally acquired data
- insecurely stored the data
- processing the data in a manner that impacts human rights like freedom of expression[70]
- using the data in a way that violates the relevant privacy and user agreement or local laws
- providing the data to a customer or other business partner that then uses it to adversely affect human rights
- failing to address such a human rights impact

Businesses should anticipate each of these grievances and establish an effective and accessible mechanism through which they can be raised and addressed.

[70] Facebook, for example, was recently accused of processing news stories in a manner that promoted liberal news and suppressed conservative news within its platform. See, e.g., Dave Lee, "Facebook: Political bias claim 'untrue'", *BBC News Services* (May 10, 2016), available at: http://www.bbc.com/news/technology-36254201. Because particular viewpoints were allegedly suppressed, this could constitute an infringement of the freedom of expression.

7. Recommendations

The need for change is apparent, most starkly due to the lack of transparency and accountability surrounding big data. Experts have outlined several proposals as to how this can be brought about, which encompass both regulatory solutions and corporate action, including:

Increase Transparency

The first step towards eliminating the shadows that shroud current big data analytical practices is to increase transparency. This is critical to establishing effective legislation and regulation, evaluating and addressing any gaps, and ensuring remedies are effective and accessible, as outlined in the UN Guiding Principles on Business and Human Rights. The first step in that process is expanding user knowledge about how data is being generated, gathered, sold, and used. The next is changing the regulatory focus from data collection to data use so that the regulatory system is clearly understood and transparent to users.

Expand User Knowledge

Users need to know far more than they currently do to effectively evaluate policies, agreements, and otherwise protect their data. In short:

> *"people should be entitled to know what data is being collected about them, what data is being archived about them, and how data about them is being used – and by whom."*[71]

If the data is being collected via an app, third-party cookie, or other form of electronic tracking, then the manufacturer or web host should disclose what specific information is being collected and the specific parties to which it is being distributed, sold, or otherwise shared.[72] If that data is being fed into an algorithm that affects how an individual is evaluated for credit or some other purpose, then the algorithm should be disclosed or otherwise explained so users understand – and have some measure of control over – the components being considered.[73] If the data is being stored, users are entitled to know where it is being stored and the level of cybersecurity that will be provided.[74] This information should be clearly delineated within any user agreement or privacy policy so that the average user can easily access and understand the information and make an informed decision.[75]

A notice and consent framework would be a natural fit for accomplishing this goal, but because this already exists in the current framework and contains acknowledged flaws, it would likely need to be updated with standardised policies and a third-party certification process that could be easily understood and compared by users.[76] Several privacy specialists have recommended establishing an independent regulatory agency dedicated to accomplishing this and otherwise

[71] Schneier, *supra* note 2, at 159

[72] *Id.*

[73] *Id.* at 160. This creates problems when businesses claim that their algorithms are trade secrets, but the direct impact on the public arguably trumps this proprietary interest and there are ways to audit such algorithms for fairness without releasing them to the public or otherwise endangering their proprietary nature. *Id.* at 196.

[74] *Id.* at 159.

[75] *Id.*

[76] *Id.* at 204.

protecting privacy,[77] while several privacy-oriented companies and organisations have developed privacy-enhancing technologies to use in the interim. Disconnect, for example, is rolling out a tool called "Privacy Icons" that translates the privacy policies of numerous websites into a legend of emblems that users can easily understand.[78]

Increase User Control

Another way to eliminate the shadows that shroud current big data analytical practices is by increasing user control over data by creating a more meaningful, nuanced form of user consent – such as the ones discussed in the personal privacy profiles and data wrappers described above – or creating an opt-in rather than opt-out system.[79] Whatever the method chosen, the solution should include a consent-based system that "*recognise[es] that people can engage in privacy self-management only selectively*", "*adjust[s] privacy law's timing to focus on downstream uses*", and includes "*a coherent approach to consent … that accounts for the social science discoveries about how people make decisions about personal data*".[80] It should also include a right to delete whenever a user breaks ties with a company or objects to a downstream use.[81]

Transparency is central to user control of data. Users cannot exercise the proposed privacy rights or provide meaningful consent if they cannot decipher the various privacy policies and user agreements that govern data use. These recommendations are dependent on expanded user knowledge and data-use-based regulation.

Regulate Data Use Rather than Data Collection

As important as it is to expand user knowledge of how data is generated, collected, sold, and used so that users can effectively evaluate policies and agreements, it is critical to develop a consent-based system that recognises that "*people can engage in privacy self-management only selectively*."[82] Big data analytics have become such a common practice that it is increasingly unreasonable to expect that an individual can honestly evaluate each of these policies and agreements:

> *"[T]here is simply so much data being collected, in so many ways, that it is practically impossible to give people a meaningful way to keep track of all the information about them that exists out there, much less to consent to its collection in the first place."*[83]

The solution is to shift the regulatory focus from the point of collection to the point of use.[84] Indeed, users tend to equate privacy violations with data use abuses rather than the data collection itself:

[77] *See supra* discussion in Section D.1.
[78] Disconnect, https://disconnect.me/icons (last visited April 24, 2016).
[79] Schneier, *supra* note 2, at 198
[80] Solove, *supra* note 50, at 1903.
[81] Schneier, *supra* note 2, at 201-02 (We should be able to tell any company we're entrusting our data to, "*I'm leaving. Delete all the data you have on me.*" We should be able to go to any data broker and say, "*I'm not your product. I never gave you permission to gather information on me and sell it to others. I want out of your database.*").
[82] Solove, *supra* note 50, at 1903.
[83] Mundie, *supra* note 25.
[84] *Id.*

> *When people are asked to give a practical example of how their privacy might be violated, they rarely talk about the information that is being collected. Instead, they talk about what might be done with that information, and the consequences: identity theft or impersonation, personal embarrassment, or companies making uncomfortable and unwelcome inferences about their preferences or behavior. When it comes to privacy, the data rarely matters, but the use always does.*[85]

But users, who must currently give their consent at the point of collection, receive little to no specific information within those agreements about how that data will be used.[86] Moreover, that original notice and consent does nothing to limit the down-stream uses of that data if the agreement permits data sales to third parties. Users rationally fear that they "*do not know who possesses data related to them and have no way to know whether the information is being used in acceptable ways*."[87] To be effective, data privacy regulation must shift from requiring consent at the point of data collection to creating a consent framework around data use. This also permits businesses to use information that users ceded for a particular purpose – such as preventing unauthorised access to an account – while also honoring users' objections to it being used in unintended ways.[88]

One suggested mechanism for accomplishing this is creating data "wrappers" that describe the type of material it contained (without revealing content) and include rules around how and when that data can be accessed and used.[89] The wrappers would essentially act as a virtual "lock" against unauthorised use.[90] A regulatory agency would hold this "key", distribute it only according to user instructions, and conduct compliance audits of any business that accessed such data.[91] According to PCAST, which incorporated this idea and expanded upon it in its presidential report, these wrapper-instructions could be created based on a user-created personal privacy profile.[92] The profile would contain nuanced instructions regarding how a user's data could be used after collection and then be translated into code, rendered tamper-proof, and attached to all data associated with that person.[93]

Regulate Data Brokers

Finally, it has been suggested that data brokers should be regulated. Because the current data privacy framework focuses on user consent at the point of collection rather than data use, there is a gap in which the data broker business operates with impunity. It essentially creates an exception in which data brokers – because they do not collect data directly from a user, and the majority of agreements do not govern downstream uses – are not regulated. This exception needs to be eliminated. Data privacy regulation will only be successful if every data controller – including those that do not interact directly with the user – are included in the regulations.[94]

[85] *Id.*

[86] *Id.*

[87] *Id.*

[88] Schneier, *supra* note 2, at 196 ("*The point of use is a sensible place to regulate, because much of the information that's collected about us is collected because we want it to be. We object when that information is being used in ways we didn't intend: when it is stored, shared, sold, correlated, and used to manipulate us in some stealthy way. This means that we need restrictions on how our data can be used, especially restrictions on ways that differ from the purposes for which it was collected.*").

[89] Mundie, *supra* note 25.

[90] *Id.*

[91] *Id.*

[92] PCAST Presidential Report, *supra* note 52, at 41.

[93] *Id*

[94] This proposition aligns with UN Guiding Principle 13, which states that businesses can be responsible for both direct and indirect human rights impacts, such as those indirect impacts that result from business relationships. *See supra* note Error! Bookmark not defined. and accompanying text.

One way to begin incorporating data broker regulation is to require:

> *"both government agencies and private companies engaging in mass data collection to file Privacy Impact Notices, modeled after Environmental Impact Reports. This would serve to inform the public about what's being collected and why, and how it's being stored and used."*[95]

Another is to designate particular businesses as automatically incurring fiduciary duties – such as a particular level of security, audit requirements, and so forth – because of the "*large amount of personal information they naturally collect: ISPs, cell phone companies, e-mail providers, search engines, social networking platforms*".[96] Any requisite audits could be conducted by "algorthmists," or individuals who would "*take a vow of impartiality and confidentiality . . . [and] evaluate the selection of data sources, the choice of analytical and predictive tools, including algorithms and models, and the interpretation of results*."[97]

[95] Schneier, *supra* note 2, at 198.
[96] *Id.* at 205.
[97] Mayer-Schönberger & Cukier, *supra* note 2, at 180.

8. Conclusion

As consumers we create data about ourselves on a daily basis. That data is not always personal data, but there is no doubt that data relating to our behaviour and how to predict it is financially valuable. Data is the new currency, collected and traded, and is a huge economic driver for industry.[98] Due to the scale and variety of data being generated, implementing safeguards for privacy and other rights has become even more challenging.

The difficulties involved in balancing the risks and benefits posed by big data analytics will affect companies in all sectors, not just technology, because of the growing reliance on business models built around big data. All companies have a responsibility to respect human rights and should conduct due diligence exercises to mitigate the risks that the collection, storage, processing and sharing of data presents. But the lack of transparency and consumer awareness makes it much more difficult for individuals to hold parties accountable.

A small number of high profile companies are actively engaged in identifying the implications of big data analytics and associated business responsibilities. But many of the companies profiting from the big data revolution are not engaged in this conversation and continue to benefit from the lack of transparency and accountability.

It is easy to get bogged down with the infinite possibilities and complexities that technology and big data analytics present. But businesses cannot afford to ignore the negative impacts big data analytics can have on individuals or their responsibility to respect those individuals' human rights throughout their operations. This paper provides several insights and recommendations into how businesses, and data brokers in particular, should begin to address such concerns.

[98] See generally Joris Toonders, "Data Is the New Oil of the Digital Economy," *Wired*, (last visited 27 July 2016), available at: https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/