THE CENTRE FOR internet & society

Benefits, Harms, Rights, and Regulation: A Survey of Literature on Big Data

EDITORS **Sunil Abraham, Elonnai Hickok, Leilah Elmokadem**
AUTHORS **Amber Sinha, Vanya Rakesh, Vidushi Marda, Geethanjali Jujjavarapu**

The Centre for Internet and Society, India

This work has been made possible by a grant from the John D. and Catherine T. MacArthur Foundation. The conclusions, opinions, or points of view expressed in the report are those of the authors and do not necessarily represent the views of the John D. and Catherine T. MacArthur Foundation.

Designed by **Saumyaa Naidu**

Contents

1. Introduction

In 2011, it was estimated that the quantity of data produced globally surpassed 1.8 zettabyte.[1] By 2013 it had increased to 4 zettabytes.[2] With the nascent development of the so-called 'Internet of Things' gathering pace, these trends are likely to continue. This expansion in the volume, velocity, and variety of data available[3], together with the development of innovative forms of statistical analytics, is generally referred to as "Big Data"; though there is no single agreed upon definition of the term. Although still in its initial stages, big data promises to provide new insights and solutions across a wide range of sectors, many of which would have been unimaginable even a decade ago.

Despite enormous optimism about the scope and variety of big data's potential applications, many remain concerned about its widespread adoption, with some scholars suggesting it could generate as many harms as benefits.[4] Most notably are the concerns about the inevitable threats to privacy associated with the generation, collection and use of large quantities of data.[5] Concerns have also been raised regarding, for example, the lack of transparency around the design of algorithms used to process the data, over-reliance on big data analytics as opposed to traditional forms of analysis and the creation of new digital divides.

The existing literature on big data is vast. However, many of the benefits and harms identified by researchers tend to focus on sector specific applications of Big Data analytics, such as predictive policing, or targeted marketing. Whilst these examples can be useful in demonstrating the diversity of big data's possible applications, they do not offer a holistic perspective of the broader impacts of Big Data.

This survey draws upon a range of literature including news articles, academic articles, and presentations and seeks to disaggregate the potential benefits and harms of big data, organising them into several broad categories that reflect the existing scholarly literature. The survey also recognises the non-technical big data regulatory options which are in place as well as those which have been proposed by various governments, civil society groups and academics.

2. Benefits

In recent years, scholars, politicians and business leaders have confidently proclaimed big data as a potential solution to a diverse range of problems from, world hunger and diseases to government budget deficits and corruption. According to a report by the McKinsey Global institute[6], much of modern economic activity would not take place without big data. But if we look beyond the hyperbole and headlines, what do we really know about the advantages of big data?. Instead of trying (and failing) to list the many possible applications of big data analytics across all sectors and industries, for the purposes of this survey we have attempted to distil the various advantages of Big Data discussed within literature into the following five broad categories; Decision-Making, Efficiency & Productivity, Research & Development, Personalisation and Transparency, each of which will be discussed separately below.

2.1 Decision-Making

Whilst data analytics have always been used to improve the quality and efficiency of decision-making processes, the advent of big data means that the areas of our lives in which data-driven decision-making plays a role is expanding dramatically as businesses and governments become better able to exploit new data flows. Furthermore, the real-time and predictive nature of decision-making made possible by big data are increasingly allowing these decisions to be automated. As a result, big data is providing governments and business with unprecedented opportunities to create new insights and solutions, becoming more responsive to new opportunities and better able to act quickly - and in some cases preemptively - to deal with emerging threats. Depending on the purpose for which one needs the data in question, big data analytics can be broadly categorised into two brackets - descriptive and predictive. Both descriptive and predictive big data are useful. The former provides counters for events - number of posts, vehicles and so on, which can be used to build upon existing databases (such as traffic and so on). The latter involves modelling the data collected to make predictions and analyses of trends and the future (such as the kind of products most customers would be interested in buying, monitoring oil and gas pipelines and so on). Both categories allow for better decision making processes across sectors.

This ability of big data to speed up and improve decision-making processes can be applied across all sectors from transport to healthcare and is often cited within the literature as one of its key advantages. Joh, for example, highlights the increased use of data driven predictive analysis by police forces to help them to forecast the times and geographical locations in which crimes are most likely to occur. This allows the force to redistribute their officers and resources according to anticipated need, and in certain cities, has been highly effective in reducing crime rates.[7] Raghupathi meanwhile cites the case of healthcare, where predictive modelling driven by big data is being used to proactively identify patients who could benefit from preventative care or lifestyle changes.[8] In the United Kingdom, the government makes healthcare outcomes public, so that citizens can make informed decisions on providers.[9]

Big data can also help go a long way in creating values for industries and businesses by enabling economic surplus.[10] Services like information technology, insurance, products like computers, and governance are all poised to receive a boost from the use of big data.

One area in particular where the decision-making capabilities of big data are having a significant impact is in the field of risk management.[11] For instance, big data can allow companies to map their entire data landscape to help detect sensitive information, such as 16 digit numbers - potentially credit card data - which are not being stored according to regulatory requirements and intervene accordingly. Similarly, detailed analysis of data held about suppliers and customers can help companies identify those in financial trouble, allowing them to act quickly to minimize their exposure to any potential default.[12]

2.2 Efficiency and Productivity

In an era when many governments and businesses are facing enormous pressures on their budgets, the desire to reduce waste and inefficiency is prominent . By providing the information and analysis needed for organisations to better manage and coordinate their operations, big data can help alleviate such problems, leading to better utilization of scarce resources and a more productive workforce.[13]

Within the literature, such efficiency savings are most commonly discussed in relation to reductions in energy consumption.[14] For example, a report published by Cisco notes how the city of Oslo has managed to reduce the energy consumption of street-lighting by 62 percent through the use of smart solutions driven by big data.[15] Increasingly, however, statistical models generated by big data analytics are also being utilized to identify potential efficiencies in sourcing, scheduling and routing in a wide range of sectors from agriculture to transport. For example, Newell observes how many local governments are generating large databases of scanned license plates through the use of automated license plate recognition systems (ALPR), which government agencies can then use to help improve local traffic management and ease congestion.[16] Research also indicates that through the use of big data analytics, Europe's public sector could reduce the cost of administration by anywhere between 15-20%, creating between $223-446 billion in value through efficiency gains as well as systematic tax collection.[17]

Commonly these efficiency savings are only made possible by the often counter-intuitive insights generated by the big data models. For example, whilst a human analyst planning a truck route would always tend to avoid 'drive-bys' - bypassing one stop to reach a third before doubling back - big data insights can sometimes show such routes to be more efficient. In such cases, efficiency saving of this kind would in all likelihood have gone unrecognised by a human analyst who is not trained to look for such patterns.[18] Thus, digital collection of large quantities of data can allow organisations and governments to collect more accurate and detailed information on multiple indicators, and therefore expose variability to boost performance.[19]

2.3 Research, Development, and Innovation

Perhaps one of the most intriguing benefits of big data is its potential utility in the research and development of new products and services. As is highlighted throughout the literature, big data can help business to gain an understanding of how others perceive their products or identify customer demand and adapt their marketing and products accordingly.[20] Analysis of social media data, for instance, can provide valuable insights into customers' sentiments towards existing products as well as discover demands for new products and services, allowing businesses to respond more quickly to changes in customer behaviour.[21]

In the pharmaceutical industry, the use of big data could result in value creation of around $100 billion. The use of electronic data capture could help in recording patient information in the provider's electronic medical records, which could become a primary source for clinical-trial data, reducing the likelihood of data errors caused by manual or duplicate entry. Remote monitoring of sites with real time data access could improve management and responses to issues that arise during trials, improving clinical trial efficiency.[22]

Similarly, in the field of computational social sciences, big data has facilitated many types of nuanced analysis. Social and electronic media sources are being relied upon by social scientists to cast insight on issues like subjective well-being, scoring of the emotionality of news content, understanding dynamics between conventional and social media during presidential debates. [23]

In addition to market research, big data can also be used during the design and development stage of new products. by helping to test thousands of different variations of computer-aided designs in an expedient and cost-effective manner. In doing so, business and designers are able to better assess how minor changes to a products' design may affect its cost and performance, thereby improving the cost-effectiveness of the production process and increasing profitability.

2.4 Personalisation

For many consumers, perhaps the most familiar application of big data is its ability to help tailor products and services to meet their individual preferences. This phenomena is most immediately noticeable on many online services such as Netflix; where data about users activities and preferences is collated and analysed to provide a personalised service, for example by suggesting films or television shows the user may enjoy based upon their previous viewing history.[24] By enabling companies to generate in-depth profiles of their customers, big data allows businesses to move past the 'one size fits all' approach to product and services design and instead quickly and cost-effectively adapt their services to better meet customer demand.

In addition to service personalisation, similar profiling techniques are increasingly being utilized in sectors such as healthcare. Here, data about a patient's medical history, lifestyle, and even their gene expression patterns are collated, generating a detailed medical profile which can then be used to tailor treatments to meet their specific needs.[25] Targeted care of this sort can not only help reduce costs by helping avoid over-prescriptions, but may also help improve the effectiveness of treatments and ultimately their outcomes.

2.5 Transparency

If 'knowledge is power', then, - so say big data enthusiasts - advances in data analytics and the quantity of data available can give consumers and citizens the knowledge to hold governments and businesses to account, as well as make more informed choices about the products and services they use. Nevertheless, data (even in immense quantities) does not necessarily equal knowledge. In order for citizens and consumers to fully utilize the vast quantities of data available to them, they must first have some way to make sense of it. For some, big data analytics provides such a solution, allowing users to easily search, compare and analyze available data, thereby helping to challenge existing information asymmetries and make business and government more transparent.[26]

In the private sector, big data enthusiasts have claimed that it holds the potential to ensure complete transparency of supply chains, enabling concerned consumers to trace the source of their products. For example, this can help consumers ensure that the products they purchase have been sourced ethically.[27] Furthermore, big data is now making accessible information which was previously unavailable to the average consumer, and challenging companies whose business models rely on the maintenance of information asymmetries. The real-estate industry, for example, relies heavily upon its ability to acquire and control proprietary information, such as transaction data, as a competitive asset. In recent years, however, many online services have allowed consumers to effectively bypass agents, by providing alternative sources of real-estate data and enabling prospective buyers and sellers to communicate directly with each other[28] thereby providing consumers with access to large quantities of actionable data. Big data can help eliminate established information asymmetries, allowing consumers to make more informed decisions about the products they buy and the services they enlist.

Big data's potential to improve transparency and accountability can also be seen in the public sector, with many scholars suggesting that greater access to government data could help stem corruption and make politics more accountable. This view was recently endorsed by the UN which, in a report published by the Independent Expert Advisory Group on the "Data Revolution for Sustainable Development"[29], highlighted the potential of big data to improve policy-making and accountability. In the report, experts emphasize the potential of what they term the 'data revolution' to help achieve sustainable development goals by, for example, helping civil society groups and individuals 'develop data literacy and help communities and individuals to generate and use data, to ensure accountability and make better decisions for themselves'.[30]

Thus, big data can often lead to what has been described as the 'transparency paradox'. Big data aspires towards making the world more transparent. However, its collection is invisible, and its tools and techniques themselves are not transparent. Privacy shrouds big data analytics procedures, and thus seems to jeopardize the very outcomes that it seeks to achieve. This paradox must be acknowledged.[31]

3. Harms

The existing literature and discussions about the potential harms of big data are perhaps understandably dominated by concerns about privacy. However there is also a range of other harms which we have listed below:

3.1. Privacy

Many of the benefits outlined above exist in tension with the right to privacy. Efficient decision making, accuracy and personalisation are often achieved through widespread collection and processing of personal data, which amounts to a breach of the data subject's privacy. The following issues highlight the ways in which big data analytics compromises an individual's privacy:

3.1.1. Re-identification

Traditionally many big data enthusiasts have used de-identification [the process of anonymising data by removing Personally Identifiable Information(PII)] as a way of justifying mass collection and use of personal data. However, many scholars remain concerned about the limits of anonymisation. As Tene & Polonetsky observe, 'Once data, such as a clickstream or a cookie number, are linked to an identified individual, they become difficult to disentangle'.[32] They cite the example of University of Texas researchers Narayanan and Shmatikov, who were able to successfully re-identify anonymised Netflix user data by cross referencing it with data stored in a publicly accessible online database. As Narayanan and Shmatikov explained,"once any piece of data has been linked to a person's real identity, any association between this data and a virtual identity breaks anonymity of the latter".[33] The quantity and variety of datasets which big data analytics has made associable with individuals is therefore expanding the scope of the types of data that can be considered PII, undermining claims that de-identification alone is sufficient to ensure privacy for users.

3.1.2. Privacy Frameworks Obsolete

In recent decades, privacy and data protection frameworks based upon a number of 'privacy principles'[34] have formed the basis of most attempts to encourage greater consideration of privacy issues online.[35] For many, however, the emergence of big data has raised questions about the extent to which these data protection principles are workable in an era of ubiquitous data collection.

1. ***Collection Limitation and Data Minimization* :** Big Data seeks to gather as much data as possible, instead of utilizing sampling techniques in order to understand the phenomenon being studied. This goal of attaining comprehensive coverage exists in tension with the key privacy principles of collection limitation and data minimization, which seek to limit both the quantity and variety of data collected about an individual to the absolute minimum.[36]
2. ***Purpose Limitation:*** Since the utility of a given dataset is often not easily identifiable at the time of collection, datasets are increasingly being processed several times for a variety of different purposes. Such practices have significant implications for the principle of purpose limitation, which aims to ensure that organizations are open about their reasons for collecting data, and that they use and process the data for no other purpose than those initially specified.[37]
3. ***Notice:*** As data streams become more complex, some have begun to question suitability of consent as a mechanism to protect privacy. In particular, commentators have noted that, given the complexity of data flows in the digital ecosystem, most individuals are not well placed to make truly informed decisions about the management of their data.[38] In one study, researchers demonstrated how by creating the perceptions of control, users were more likely to share their personal information, regardless of whether or not the users had actually gained control.[39] As such, the garnering of consent is increasingly becoming a symbolic box-ticking exercise which achieves little more than to irritate and inconvenience customers whilst providing a burden for companies and a hindrance to growth and innovation.[40]
4. ***Access and Correction:*** The use of big data technology by firms and companies has resulted in a continuous flow of personal data, due to which adjustment and correction of the data by the data subjects has become difficult. This creates a number of potential compliance problems such as difficulty in erasing, retrieving or correcting data. A typical big data system is not built for interactivity, but for batch processing. This also makes the application of changes on a (presumably) static data set difficult'.[41]
5. ***Choice and Consent:*** The notion that the provision of data should be a matter of personal choice for the individual and that they can, if they wish, decide to 'opt-out' of data collection (for example by ceasing use of a particular service), is an important component of privacy and data protection frameworks. However, the proliferation of internet-enabled devices and their integration into the built environment, as well as the real-time nature of data collection and analysis, are beginning to undermine this concept. For many critics of big data, compulsory provision of data as a prerequisite for the access and use of many key online services is making opting-out of data collection not only impractical but in some cases impossible.[42]
6. ***Mosaic Effect:*** Mosaic effect is where different factors can be used in conjunction with one another to determine if an individual is identifiable. This approach can enable identification even when traditional forms of PI data are protected.[43]

3.1.3. "Chilling Effects"

The normalization of large scale data collection is steadily producing a widespread perception of constant surveillance amongst users. Foucault's analysis of Jeremy Bentham's panopticon and the disciplinary effects of surveillance can be used to argue that this perception of permanent visibility can cause users to sub-consciously 'discipline' and self- regulate their own behavior, fearful of being targeted or identified as 'abnormal'.[44] As a result, the pervasive nature of big data prevents individuals from voicing and expressing their opinions, thus generating a 'chilling effect' on user behavior and free speech.

One study, conducted by researchers at MIT (Massachusetts Institute of Technology) sought to assess the impact of Edward Snowden's revelations about NSA surveillance programs on Google search trends. Nearly 6,000 participants were asked to individually rate certain keywords for their perceived degree of privacy sensitivity along multiple dimensions. Using Google's own publicly available search data, the researchers then analyzed search patterns for these terms before and after the Snowden revelations. In doing so they were able to demonstrate a reduction of around 2.2% in searchers for those terms deemed to be most sensitive in nature. According to the researchers themselves, the results 'suggest that there is a chilling effect on search behaviour from government surveillance on the Internet'.[45] Although this study focussed on the effects on government surveillance, for many privacy advocates, the growing pervasiveness of big data risks generating similar results.[46]

3.1.4. Dignitary Harms of Predictive Decision-Making

The automated nature of big data analytics possesses the potential to inflict 'dignitary harms' on individuals, by revealing insights about themselves that they would have preferred to keep private.[47] In an often cited example, following a shopping trip to the retail chain called Target, a young girl began to receive mail at her father's house advertising products for babies including, diapers, clothing, and cribs. In response, her father complained to the management of the company, incensed by what he perceived to be the company's attempts to "encourage" pregnancy in teens. A few days later however, the father was forced to contact the store again to apologise, after his daughter had confessed to him that she was indeed pregnant. It was later revealed that Target regularly analyzed the sale of key products such as supplements or unscented lotions in order to generate "pregnancy prediction" scores, which could be used to assess the likelihood that a customer was pregnant and to therefore target them with relevant offers.[48] Such cases, though anecdotal, illustrate how big data, if not adopted sensitively, can lead to private and/or personal information about users being made public.

3.2 Discrimination

3.2.1 Discriminatory Outcomes

Big data technologies provide innumerable opportunities to increase efficiency and mitigate risk. By removing human intervention and oversight from the decision-making process, firms, governments and companies using this technology run the risk of becoming blind to unfair or injudicious results generated by conscious discrimination and discriminatory programming of the algorithms. This can be categorized into Disparate Treatment, which is a conscious discrimination by the analysts and Disparate Impact, where there is no intentional discrimination, but the bias is inherent in the algorithm producing the results.[49]

Disparate Treatment

Conscious discrimination occurs when the individual is discriminated against by the data controller based on certain protected or behavioural characteristics after analysing the data produced by an algorithm. For example, it is noted that big data is increasingly being used to evaluate applicants for entry-level service jobs. One method of evaluating applicants is by the length of their commute - the rationale being that employees with shorter commutes are statistically more likely to remain in the job longer. However, since most service jobs are typically located in town centers while poorer neighborhoods are often on the outskirts of town, such criteria can have the effect of unfairly disadvantaging those living in economically underprivileged areas. Consequently, such metrics of evaluation can therefore also unintentionally act to reinforce existing social inequalities by making it more difficult for economically disadvantaged communities to work their way out of poverty.[50]

Disparate Impact

Discrimination can also be inherent in the algorithm that is processing the data and therefore lead to biased decisions or outcomes, resulting in certain groups being excluded or deprived of opportunities and benefits.[51] There currently exists a large number of automated decision-making algorithms in operation across a broad range of sectors including, most notably perhaps, those used to assess an individual's suitability for insurance or credit. In either of these cases, faults in the programming or discriminatory assessment criteria can have potentially damaging implications for the individual, who may as a result be unable to attain credit or insurance. This concern with the potentially discriminatory aspects of big data is prevalent throughout the existing literature. Real life examples have been identified by researchers in a large number of major sectors in which big data is currently being used.[52] Yu, cites the case of the insurance company Progressive, which required its customers to install 'Snapsnot'- a small monitoring device - into their cars in order to receive their best rates. The device tracked and reported the customer's driving habits, and offered discounts to those drivers who drove infrequently, broke smoothly, and avoided driving at night - behaviors that correlate with a lower risk of future accidents. Although this form of price differentiation provided incentives for customers to drive more carefully, it also had the unintended consequence of unfairly penalizing late-night shift workers. As Yu observes, 'for late night shift-workers, who are disproportionately poorer and from minority groups, this differential pricing provides no benefit at all. It categorizes them as similar to late-night party-goers, forcing them to carry more of the cost of the intoxicated and other irresponsible driving that happens disproportionately at night'.[53]

3.2.2. Inequality leading to exclusion and marginalisation

While Big Data has the potential to contribute to economic development and generate new innovations, some critics remain concerned about how the benefits of big data will be distributed and how the plausible unequal distribution would affect already established digital divides.[54] Although the power of big data is already being utilized effectively by most economically developed nations, the same cannot be said for many developing countries. A combination of lower levels of connectivity, poor information infrastructure, underinvestment in information technologies and a lack of skills and trained personnel, make it far more difficult for some global south countries to fully reap the rewards of big data. In India, there is huge challenge due to the issue of dark data. Only 16% individuals leave a digital footprint.[55] Large parts of the population is unrepresented in digital data. Further, the veracity and accuracy of the data available remains a big issue because of the complexities of local contexts - making it hard to devise a classification system which is accurate. As a consequence, the Big Data revolution risks deepening global economic inequality as developing countries find themselves unable to compete with data-rich nations whose governments can more easily exploit the vast quantities of information generated by their technically literate and connected citizens.

Likewise, to the extent that the big data analytics is playing a greater role in public policy-making, the capacity of individuals to generate large quantities of data could potentially have an impact on the extent to which they can provide inputs into the policy-making process. In a country such as India for example, where there are high levels of inequality in access to information and communication technologies and the internet, there remain large discrepancies in the quantities of data produced by individuals. As a result there is a risk that those who lack access to the means of producing data will be disenfranchised, as policy-making processes become configured to accommodate the needs and interests of a privilege minority.[56] It has also been claimed that far from helping to alleviate inequalities, the advent of big data risks exacerbating already significant digital divides that exist as well as creating new ones.[57] Also, the consequences of exclusion could be highly profound as those left out of the big data revolution may suffer tangible economic harms, for example, in case of businesses ignoring or undervaluing the preferences and behaviors of consumers who do not shop in ways that big data tools can easily capture, aggregate, and analyze. Also, the governments may come to rely on big data to such a degree that exclusion from data flows may lead to exclusion from civic and political life[58], for example in case of Aadhaar in India, whereby people are 'locked out' of public services for not having or having Aadhaar number with incorrect demographic details[59] and are excluded from welfare services when the Aadhaar-based authentication process fails.[60]

3.2.3 Event Size Bias

The nature of data impacts what bias may be built into it. Data available to business and industry is perfect and usually complete, whereas data that focusses on social sciences, human rights etc are more often than not, skewed. One of the main ways this happens is through event size bias - where larger events and issues are reported more, but smaller ones may not be reported at all.[61] This difference in the observed sample and underlying population of interest brings about a discrepancy in conclusions, decisions and accountability. These changes are brought about in the documentation process, not in the actual occurrence of events. To adjust for these biases and also other invisible biases, one can speculate and recognize bias but that would probably lead to more questions than answers.

3.2.4 Lack of Algorithmic Transparency

For many companies, the quality of their algorithms is often a crucial factor in providing them with a market advantage over their competitor. Given their importance, the secrets behind the programming of algorithms are often closely guarded by companies, and are typically classified as trade secrets and as such are protected by intellectual property rights. Although companies may claim that such secrecy is necessary to encourage market competition and innovation, many scholars are becoming increasingly concerned about the lack of transparency surrounding the design of these most crucial tools.

In particular, there is a growing common sentiment amongst many researchers that there currently exists a chronic lack of accountability and transparency in terms of how big data algorithms are programmed and what criteria are used to determine outcomes.[62] As Frank Pasquale observed,

> *"Hidden algorithms can make (or ruin) reputations, decide the destiny of entrepreneurs, or even devastate an entire economy. Shrouded in secrecy and complexity, decisions at major Silicon Valley and Wall Street firms were long assumed to be neutral and technical. But leaks, whistleblowers, and legal disputes have shed new light on automated judgment. Self-serving and reckless behavior is surprisingly common, and easy to hide in code protected by legal and real secrecy".*[63]

As such, without increased transparency in algorithmic design, instances of big data discrimination may go unnoticed as analysts are unable to access the information necessary to identify them.

3.3 Anti-Competitive Practices

As a result of the reluctance of large companies to share their data, there is an increasing divide in the possession of data between small start-ups companies and their larger and more established competitors. The value-proposition of most internet and ICT companies relying on big data technologies is based on the Network Effect, which is creating monopolies or competitive oligopolies in the market.[64] Today, since the performance of many online services are often intimately connected with the collation and use of user's data, some researchers have suggested that this inequity in possession of data could lead to a reduced choice for the customers.[65] This is reducing the flexibility for the customers by making it difficult for them to shift between service providers. Further, the new market players are often discouraged to enter the market structure as they find it difficult to prosper in the absence of user data and information.

As a result, researchers including Nathan Newman of New York University have called for a reassessment and reorientation of antitrust investigations and regulatory approaches more generally 'to focus on how control of personal data by corporations can entrench monopoly power and harm consumer welfare in an economy shaped increasingly by the power of "big data".[66] Similarly a report produced by the European Data Protection Supervisor concluded that 'the scope for abuse of market dominance and harm to the consumer through refusal of access to personal information and opaque or misleading privacy policies may justify a new concept of consumer harm for competition enforcement in digital economy'.[67]

3.4 Security

In relation to cybersecurity, big data can be viewed to a certain extent as a double-edged sword. On the one hand, the unique capabilities of big data analytics can provide organizations with new and innovative methods of enhancing their cybersecurity systems. On the other, however, the sheer quantity and diversity of data emanating from a variety of sources creates its own security risks.

3.4.1. "Honey-Pot"

The larger the quantities of confidential information stored by companies on their databases the more attractive those databases may appear to potential hackers. Other security concerns associated with the centralized storage of large amounts of data include user control, user access, and third party access.[68]

3.4.2 Data Redundancy and Dispersion

Inherent to big data systems is the duplication of data to many locations in order to optimize query processing. Data is dispersed across a wide range of data repositories in different servers, in different parts of the world. As a result, it may be difficult for organizations to accurately locate and secure all items of personal information[69].

3.5 Epistemological and Methodological Implications

3.5.1 End of Theory

In 2008, Chris Anderson predicted the 'end of theory'. According to him, the 'deluge of data' created by big data would render scientific methods of hypothesis, sampling and testing obsolete. Petabytes would allow us to rely simply on correlation, and "..let statistical algorithms find patterns where science cannot".[70] This is echoed by Cukier and Mayer-Schoenberger, although they don't go so far. For them, Big Data "represents a move away from always trying to understand the deeper reasons behind how the world works to simply learning about an association among phenomena and using that to get things done."[71]

3.5.2 Abandonment of Causation

Moving from causation to correlation is, for some scholars, a hollowing out of the scientific method and an abandonment of causal knowledge in favour of shallow correlative analysis. They warn against what they see as the uncritical adoption of big data in public policy-making.[72] Fricke has also warned that this turn away from theory and studying big data without taking contemporary philosophy of science into account is unwise.[73]

3.5.3 The Hybrid Approach

Rob Kitchin advocates that it is possible to reap the benefits of big data without compromising scientific rigour or the pursuit of causal explanations. This is through a hybrid approach which utilises the combined advantages of inductive, deductive and so-called 'abductive' reasoning, to develop theories and hypotheses directly from the data.[74] With advances in computational and statistical methods as well as innovations in data visualization and methods of linking datasets, scientists can now utilise the data available to its full potential, as professor Gary King quipped '*Big Data is nothing compared to a big algorithm*'.[75] As Patrick W. Gross, commented: 'in practice, the theory and the data reinforce each other. It's not a question of data correlations versus theory. The use of data for correlations allows one to test theories and refine them'.[76]

Big data's epistemic status and various arguments around it can be understood through the work of computer scientist Jim Gray, who adopted and developed Kuhn's concept of the paradigm shift, charting history of science through the evolution of four broad paradigms: experimental science, theoretical science, computational science and exploratory science.[77] According to Gray, we are today witnessing a transition to a 'fourth paradigm of science', which he terms the *exploratory paradigm* where scientists begin with the data itself; designing programs to mine enormous databases in the search for correlations and patterns; in effect using the data to discover the rules, as opposed to developing programs based on established rules and theories as was observed in the computational paradigm.[78]

3.5.4 Data Speaks for Itself

Analysts assume that they can 'let the data speak for itself'.[79] Joh observes how big data is being used in policing and law enforcement to help make better decisions about the allocation of police resources. By looking for patterns in the crime data they are able to make accurate predictions about the localities and times in which crimes are most likely to occur and dispatch their officers accordingly.[80] For many, however, consideration of issues such as gender, race, and religious sensitivity can be just as important to good public policy-making as quantitative data; helping to contextualise the insights revealed in the data and provide more explanatory accounts of social relations.[81]

3.5.5 Questionable methodological, epistemological assumptions

Although big data analytics can be utilized to study almost any phenomena where enough data exists, many theorists have warned that simply because big data analytics *can* be used does not necessarily mean that they *should* be used.[82] Many of the claims of big data are premised upon some questionable methodological and epistemological assumptions, some of which threaten to impoverish the scientific method and undermine scientific rigour.[83] This is because interpretation is the centre of data analytics, and it is always subject to limitation and bias. Some scholars arguing for the end of theory assume that big data eliminates the need for scientific specialization. However, others point out that the abstract and intangible nature of data requires a great deal of expert knowledge and interpretive skill to comprehend. It is therefore vital that the knowledge of domain specific experts is properly utilized to help 'evaluate the inputs, guide the process, and evaluate the end products within the context of value and validity'.[84]

3.5.6 Serendipity

By starting with the data itself, some thinkers advocate that big data analysts may be able to circumvent the need for predictions or hypothesis about what one is likely to find. As Dyche observes, *'mining Big Data reveals relationships and patterns that we didn't even know to look for'*.[85] Similarly, Steadman proposes that analysts need not bother proposing a hypothesis anymore.[86] For big data enthusiasts such as Prensky, *'scientists no longer have to make educated guesses, construct hypotheses and models, and test them with data-based experiments and examples. Instead, they can mine the complete set of data for patterns that reveal effects, producing scientific conclusions without further experimentation '*.[87] In contrast, boyd and Crawford observe how big data may also lead to the practice of 'apophenia', a phenomena whereby analysts interpret patterns that do not exist, 'simply because enormous quantities of data can offer connections that radiate in all directions".[88]

3.5.7 Reframing the constitution of knowledge

These developments in statistical and computational analysis combined with the velocity variety and quantity of data available to analysts have therefore allowed scientists to pursue new types of research, generating new forms of knowledge and facilitating a radical shift in how we think about "science" itself. As boyd and Crawford note, ' *Big Data [creates] a profound change at the levels of epistemology and ethics. Big Data reframes key questions about the constitution of knowledge, the processes of research, how we should engage with information, and the nature and the categorization of reality . . . [and] stakes out new terrains of objects, methods of knowing, and definitions of social life* '.[89]

4. Rights

As explained by James Griffin, the way we understand human rights is tied intrinsically to normative agency, which is dependent on the idea of the informed self. Without sufficient information, human agency is severely compromised and human rights become of little value. In his book, On Human Rights,[90] he defines normative agency as our capacity to choose and to pursue our conception of a worthwhile life. It is important to recognise that by agency, Griffin does not mean merely the ability to perform actions. This kind of agency involves not just the conception of a 'worthwhile life' but also active autonomy. The thrust of these theories is to view human rights as protectors of human agency and autonomy. In order to be autonomous, individuals need to make real choices, and information is prerequisite for an individual to make real choices and be autonomous.

In his books, Public Opinion[91] and The Phantom Public,[92] Walter Lippmann contemplated man's ability to perceive and accurately comprehend the world. Lippmann was extremely pessimistic about the 'public' as understood in theories of democracy as sovereign and omnicompetent citizens who can make rational choices. Jonathan Obar has argued that Lippmann's arguments are vindicated in light of technologies and practices which create a 'black-box' around information and compromises the ability of the public to make autonomous choices.

Complex algorithms used in big data such as those used by Google's search engine are multi-component systems, which leads to an opacity even for programmers working on them.[93] Jenna Burrell demonstrates the futility of exercises such as code audits as the number of auditors hours that may be needed to be engaged in order to untangle the logic of the algorithms in a complicated software system, would be huge. [94] In her article on "socially consequential mechanisms of classification and ranking, such as spam filters, credit card fraud detection, search engines, news trends, market segmentation and advertising, insurance or loan qualification, and credit scoring" and other similar mechanisms that involve personal and trace data, Burrell states that more and more these mechanisms rely upon machine learning algorithms. She points out that problems of scale and complexity in the case of machine learning algorithms employed in big data technology are peculiar and distinctive. These are not characterised simply by greater number of lines of code or the number of team member or even various linkages between modules. According to Burrell, the challenge is not merely that of comprehending the code, but that being able to understand how the algorithm operates on data, in action. She argues that while it may be possible to implement machine learning algorithms in such a way that it is comprehensible, such algorithms may not be of much use. For the models to have 'accuracy of classification', they must be accompanied by a degree of inherent complexity. In his DPhil thesis, Malte Ziewitz writes about Google's search algorithm that even if you had "Larry [Page] and Sergey [Brin] at this table, they would not be able to give you a recipe for how a specific search results page comes about."[95] The route that algorithms can take to arrive a particular conclusion can be extremely circuitous.[96]

As mentioned above, machine learning algorithms build upon themselves and the internal decision logic of the algorithm evolves as it 'learns' on input data. Handling a huge number especially of heterogeneous properties of data adds complexity to the code. Machine learning techniques quickly face computational resource limits as they scale and may manage this, using techniques written into the code (such as 'principal component analysis'), which add to its opacity. While datasets may be extremely large but possible to comprehend and code may be written with clarity, the interplay between the two in the mechanism of the algorithm is what yields the complexity and thus opacity.

Increasing use of big data in different sectors such as medicine,[97] disaster management[98], finance,[99] law enforcement,[100] journalism,[101] retail[102] and education[103] will be accompanied by a decrease in our capacity to engage with the infrastructure in question. In her essay[104] and talk,[105] Saskia Sassen draws the picture of a smart, quantified city relying on big data and networked technology. She points out that in the past, great cities have always evolved through constant engagement by its residents and only, thus, have remained in in existence for so long. However, the proprietary technology being used in building new age smart cities limits the ability of the city's inhabitants to engage with and change it. According to Sassen, the answer could be the use of open source technologies which allows those interested to fiddle around with it. However, if one goes by Jenna Burrell's account, proprietary technology is only a part of the problem.[106] Even with use of open source technology and open standards, the very nature of machine learning technology endows it with extreme complexity far beyond the capacity of most to engage with.

It is also important to note that the communication channel between an individual and big data algorithms and the capacity of an individual to engage with it is extremely narrow.[107]

5. Regulation

The ubiquitous nature of Big Data analytics and questions around its conspicuousness and transparency, its growing presence in disparate sectors like medicine, credit, healthcare, law enforcement, justice system and public policy, to name a few, and the unique questions it raises about existing legal systems (or lack thereof) presents a significant regulatory challenge.[108]So far, most attempts to articulate a suitable regulatory response to Big Data are limited to an examination of data protections law, giving greater control to citizens over their data and holding companies engaged in Big Data more accountable.[109] This section seeks to discuss the existing non-technical regulatory mechanisms that have been proposed and adopted by various countries across the world.

5.1 Algorithmic Transparency and Affirmative Action

5.1.1 Algorithmic Transparency

Pasquale and others have suggested algorithmic transparency as a possible solution to data-driven discriminatory practices. Disclosing the logic and flow of data-driven decision-making can help examine the accuracy and fairness of the conclusions drawn from data. However, as acknowledged by Pasquale himself, "information flows can be very complex," and impossible for individuals to keep track of all these 'galaxies' of data.[110]Cynthia Dwork and Deirdre Mulligan have argued that "Exposing the datasets and algorithms of big data analysis to scrutiny—transparency solutions—may improve individual comprehension, but given the independent (sometimes intended) complexity of algorithms, it is unreasonable to expect transparency alone to root out bias." Citron has also suggested that agencies should be required to regularly test their system's software for bias and other errors.

However, this technique of transparency is proving to be difficult to achieve due to a growing trend of clauses in cross-border agreements that prevent signatory countries from mandating source code disclosure as a condition for import, distribution, sale or use of such software. Currently, Trans-Pacific Partnership, Trade in Services Agreement (TISA) and Regional Comprehensive Economic Partnership (R-CEP) are some of the agreements which have such clauses.

5.1.2 Affirmative Action

Anupam Chander proposes affirmative action as the remedy against algorithmic discrimination, as they replicate or amplify real-world biases through their statistical methodologies. This would involve both design choices which address specific discriminatory impacts of an algorithm, as well as focus on data which the algorithms use without needing transparency in the design of the algorithm. This method also involves evaluation of data which is used to train an algorithm for being embedded with implicit and institutional biases.[111]

5.2 Due process

5.2.1 Data Due Process

Crawford (2014) and Citron (2014) call for a data due process that would regulate the fairness of big data's analytical processes with regards to how they use personal data (or metadata derived from or associated with personal data) in any adjudicative process, including those whereby big data is being used to determine attributes or categories of an individual. The elements from traditional due process that are most suitable to data due process would be: "an unbiased tribunal", "the right to know the evidence against one", "the making of a record" and "a statement of reasons".[112]

5.2.2 Separation of Powers

Crawford (2014) highlights that due process' historical role as a means of separating powers among different branches of the government is another favourable reason to consider it as a mechanism to address how big data handles personal information. Data due process would ensure that entities with vested interests in certain outcomes of trial would not intervene with the judgement made with regards to the individual's case (denial of a service, targeting by police, etc). The neutrality of the arbiter in the data due process ensures that the individual is heard as they attempt to correct or justify the data by which they have been adjudicated or discriminated against.[113]

5.2.3 Opportunity to be Heard

Citron posits that an "opportunity to be heard" in a big data context would involve access to an automated program's source code or a hearing on the logic of computer program's decision. This opportunity has been crystallized by the European Union's GDPR (General Data Protection Regulation), where the data controller is now under an obligation to ensure that the data subject can obtain human intervention, express his/her point of view and challenge the decisions made by the algorithms.[114] To better adapt due process application to automated systems, Citron suggests that instead of subjecting every automated system to cross-examination, one could at least invest in extra education about the biases and fallacies of automation for government personnel who use the systems to make administrative decisions.[115]

5.2.4 Audit trails and testing of scoring systems

Citron suggests that agencies should require hearing officers to explain in detail their reliance on an automated system's decision, including any computer-generated facts or legal findings. She critiques automated administrative systems because they often fail to retain any audit record of how they made the decisions at issue or upon what data the decision was based. She hereby calls for an audit trail because it provides reassurance and increases accuracy. Access to these audit trails would also allow individuals to raise specific questions and objections to how and when their data is being used in various processes. Scoring systems should be subject to licensing and audit requirements when they enter critical settings like employment, insurance and health care.[116]

5.3 Data Protection Law

The European Network and Information Security Information (ENISA) is advocating for the enhancement of privacy by design in an era of big data through the implementation of mechanisms such as sticky policies, opt-out tools and local anonymization.[117]

5.3.1 Sticky Policies

These are policies that allow the data subject to decide on a set of conditions and constraints which unambiguously lay down how her/his PII is to be used by the party receiving the data.[118] As the data moves across multiple parties, these policies define an allowed usage and obligations, thus enhancing the control of the data owners over their personal information. They impose prohibitions and obligations such as access of third parties and the purpose for which the data is being used. These policies also allow the data owners to blacklist certain parties from gaining access to their personal information along with laying down rules such as a notice of disclosure and the deletion or minimization of data after a specified period of time.[119]

5.3.2 Opt-Out Tools

This mechanism provides the data subjects with a choice to opt-out of the data collection or data processing activity. The opt-out tool offers the individual with an option to avoid his/her data from being collected. This also extends to the data processing level, where the individual can at any point opt-out and prevent his data from being processed for any further purpose or use.[120]

5.3.3 Local Anonymization

This would enable an individual or the receiver who is processing the data of the individual to weed out the personal information of the data subject before divulging the data for analytical purposes. This solution proves to be useful in cases where personal data is not relevant for the analysis and the collection of anonymized data is sufficient.[121]

5.3.4 Re-identification as an offence

Preventing re-identification of an individual's personal data after de-identification helps in maintaining anonymity and secures the privacy of the data owner. The Australian Privacy Amendment (Re-identification offence) Bill provides for preventing re-identification of information and mandates non-disclosure of personal information if re-identified.[122]

5.3.5 Anti-Profiling

It has been suggested and subsequently adopted that automated processing should be subject to human intervention and explicit consent by the data subject.[123] The data subjects also have a right to object to processing of their personal data carried on for particular purposes.[124] If the automated profiling of the individuals is having some legal effects or is significantly affecting them, they have a right to not be subjected to such profiling.[125]

5.3.6 Privacy Impact Assessments (PIAs)

This assessment includes setting out the goals and benefits of processing the information, along with the effect it will have on the data owner's' privacy.[126] This is a methodology for organisations to apply, calibrate and implement abstract privacy obligations based on the actual risks and benefits of the proposed data processing.[127] This would mean addressing the question whether "there is a significant likelihood that an identified threat could lead to a recognised harm with a significant degree of seriousness." PIAs will only be effective in the era of big data if there is a coherent understanding of privacy impact by those who have the technical expertise in applying algorithms and designing.[128]

5.3.7 Do-Not-Track Tools

Strengthening do not track tools helps the consumers have greater control over when and how their data is used considering the expanding technologies capable of tracking and recording individual behavior, patterns and actions.[129]

5.3.8 Pseudonymization

This is a privacy enhancing technique which was featured in the GDPR where the personal information is processed in such a way that it can no longer be attributed to the data subject.[130]

5.3.8 Privacy Nudges

The insights from cognitive science, particularly using the theory of nudge would be an acceptable compromise between the inefficacy of privacy self-management and the dangers of paternalism.[131] The rationale is that while nudges influence choice, they are not overly paternalistic in that they still give the individual the option of making choices contrary to those sought by the choice architecture.

5.3.9 Privacy Defaults

A rule which mandates that data collectors set optimal defaults that ensure that the most sensitive information is subjected to least degree of disclosure unless otherwise chosen by the user, will ensure greater privacy protection.

5.3.10 Privacy Through Usability

The usability approach entails designing the system in way that is most intuitive and easy for users to decide whether to provide the information, along with a soft paternalistic approach which seeks to aid the decision-making by providing other information such as how many people would have access to the information, if provided, and set defaults such that the information is not visible to others unless explicitly set by the user.

5.3.11 Privacy by design

A design approach to privacy notices which includes looking at factors such as the timing of the notice, the channels used for communicating the notices, the modality (written, audio, machine readable, visual) of the notice and whether the notice only provides information or also include choices within its framework, would be of great help. Further, use of privacy by design principles can be done not just at the level of privacy notices but at each step of the information flow, and the architecture of the system can be geared towards more privacy enhanced choices.

5.3.12 Privacy Principles

1. **Purpose Limitation**: It has been mandated that data collected for a particular purpose should not be used for other analytical purposes or transferred to a third party without the consent of the data owner.[132] Article 29 Working Party has opined that the purpose limitation principle in the big data age consists of two key components: purpose specification and compatible use. Purpose specification means that the purpose for which the data is being processed must be specific, explicit and legitimate. In order to satisfy the second component, i.e. compatible use, further processing of the data must be compatible with the original purpose for which the data was collected.[133]

2. **Notice**: If the big data analytics is leading to new or unforeseen processing of information, the data receivers or controllers should inform the individuals of the same in the form of a privacy notice. This notice will be more detailed and will provide information as to how long the personal data will be stored along with whether the data receiver plans to transmit the data outside the European Economic Area.[134]

3. **Data minimisation and retention**: Personal data collected from individuals should be used sparingly in the age of big data where data is continuously being gathered and processed and the data controllers should make sure that efficient mechanisms are in place which will minimize the usage of personal information while retaining the information for a minimum period of time which is not longer than the time required for processing.[135]

4. **Openness**: With the existence of numerous entities that collect and process data on a real-time basis, it becomes difficult for the data subject to track who is collecting his/her data and for what purpose it is being used. A common portal should be developed which lists entities, outlines their data practices and enables consumers to control the use of their personal information by various parties.[136]

5.4 Anti-Discrimination Law

5.4.1 Disparate Treatment and Impact

The FTC has proposed the use of various equal opportunity laws by data controllers and analyzers in order to prevent discriminatory results and patterns that might be drawn as a result of the voluminous amounts of data collected.[137] Few academics have also pointed out that the existing laws do not extend to the discriminatory harms arising from big data analytics.[138] They have suggested that these laws need to be modified and expanded in such a way that they remedy the disparate treatment and impact resulting from biased algorithmic decisions.

5.4.2 Plausible Impairment and Genetic Information

It has been suggested that the ADA (American with Disabilities Act) should be reformed so as to account for the possibility that employers might make employment decisions depending on the data that may foresee that a candidate will develop an impairment in near future.[139] Similarly, the GINA in US[140] and the Disability Discrimination and Other Human Rights Legislation Amendment Act 2009 in Australia[141] prevent collecting genetic data and discriminating based on the same.

5.5 Competition Law

Policymakers in the US have suggested the use of Competition Law to address the big data harms especially around mergers and other commercial arrangements between firms through the following proposals which police the invasion of privacy.[142]

5.5.1 Merger Control

The regulator should investigate whether a particular merger would result in reduced incentives for the merged firm to compete on the consumer privacy protection frontier. It would also evaluate privacy as a non-price dimension of competition and examine mergers which involves large data sets.[143] Further, a merger between two companies in possession of large databases can be estopped even if the companies do not have any form of competitive relationship.

In Europe, courts have looked at questions such as the impact of combination of databases on competition[144] and have held that in the context of merger control, data can be a relevant question if an undertaking achieves a dominant position through a merger, making it capable of gaining further market power through increased amounts of customer data.[145]

5.5.2 Purpose Limitation Covenant

The regulator should try to balance the costs and benefits of the impact on competition against consumer protection in certain situations where "conduct-distorting commerce implicates both consumer protection and competition principles."[146] If competitors mutually agree to provide an undertaking to limit the use of certain data for marketing purposes, this could override the concerns about competition being harmed.

5.5.3 Abuse of Dominant Position

Companies that have achieved or maintained monopoly power by deceiving or misleading consumers about data collection practices would be held accountable under the antitrust laws.[147] Other relevant forms of regulation attempt to address the disproportionate comparative advantages enjoyed by large companies due to their complete ownership of vast consumer datasets, throttling small market entries. The Competition Tribunal recently supported the Canadian Competition Bureau's case against the Toronto Real Estate Board (TREB) for abusing its market power by preventing brokerages from offering innovative products and services to consumers over the Internet.[148] The Tribunal ruled that the restrictions TREB imposes on its members' use and display of the data in the Toronto Multiple Listing Service system are anti-competitive, having an adverse impact on innovation, quality and the range of residential real estate brokerage services available in the Greater Toronto Area.[149]

5.5.4 Introducing Competition through Liberalization and Privatization

This mechanism distributes the power to store data by one particular entity to multiple sectors or companies. It will reduce the concentration of power with respect to collection and storage of data, thus enhancing a controlled and streamlined use of the individual's personal information for processing. In the United Kingdom, the task of providing unique identities to all the citizens along with collecting and storing biometric data has been distributed among eight service providers instead of concentrating all the data in the hands of the government. This move has created an element of competition among the service providers to collect and maintain data for analytical purposes, thus reducing the amount of data held by each provider in order to ensure that there is no unwarranted use of the individual's data which might invade his/her privacy resulting in a detrimental effect.[150]

However, it has been difficult to apply competition law to address the harms of big data as there are limited circumstances in which big data meets the four traditional criteria for being a barrier to entry or a source of sustainable competitive advantage — inimitability, rarity, value, and non-substitutability.[151] In this context the peculiar ways in which network effects, multi-homing practices and how dynamic the digital markets are, are all relevant factors which could have both positive and negative impacts on competition.

5.6 Consumer Protection Law

5.6.1 Adjustment and Accuracy of Consumer Reports

The Federal Trade Commission has laid down that Credit Reporting Agencies must take all measures and follow a reasonable procedure to ensure that the consumer reports are accurate while also providing the consumers with access to their personal information along with the means to adjust, update and rectify any errors in the information.[152]

5.6.2 Deceptive and Unfair Practices by Data Controllers

The Federal Trade Commission suggested that the companies involved in big data analytics should bear in mind the promises made to the consumers and also check if the promises are being materially violated.[153] These promises could include a promise to disclose material information to the consumers, restraining from divulging data to third parties or safeguarding consumers personal information in order to ensure that the same is not used to their detriment.[154] The Commission also pointed out that the companies must not sell out information about consumers to their customers if they know that the latter are going to put that information to a fraudulent or discriminatory use.[155]

5.6.3 Role of Intermediary Firms

It has also been suggested by scholars and academics that the firms which act as middlemen should enable consumers to opt for suitable privacy preferences or identify customers that match the privacy preferences of the consumers in order to avoid the unfair use of big data by a dominant firm or company.[156]

5.6.4 Choice between payment with money and payment with data

It has been suggested that the consumers or data owners should be given an active choice between paying for an online service directly through monetary means or indirectly by allowing data collection.[157] This choice provides them with an "exit strategy" from the data collection process and will avoid unwarranted access to the individual's personal information.

5.6.5 Expanding Technical Expertise

The civil rights groups and consumer protection agencies should widen their technical expertise by developing a plan for investigating and fixing violations of law in order to address the discriminatory effects of big data analytics.[158]

5.6.6 Including data brokers in the Data Breach Notification Laws

Proposals have been made to bring in more transparency with respect to data brokers. The consumers ought to know how and with whom their data is being shared beyond the parties with which they transact directly.[159]

5.6.7 Access to information about differential pricing

Consumers should be given access to the information about whether the prices they are offered for products are substantially varying from the prices offered to other consumers. This will help them determine the existence of discrimination and disparate impact if any.[160]

5.7 Data Portability

5.7.1 Mandates

The data portability will mandate companies and data receivers to make information accessible to the individuals in a format which is machine-readable, interoperable and portable in order to enable use and reuse of the data. The individuals would also be enabled to switch the service providers if they intend to do so.[161] Technological projects such as the Berkman Centre of Internet and Society's Project VRM (Vendor Relationship Management) which addresses issues such as vendor lock-in and tries to provide better and more equitable manners of engaging with vendors are examples of data portability. Data portability also provides for the companies to allow the data owners to reap the benefits of third party applications to analyse their personal information and draw useful inferences.[162]

Data portability can happen at two levels, the individual level and the group level. Collected and processed data should be put back in the hands of the disadvantaged or underprivileged groups in such a manner that it strengthens their capacity to use it.[163] This ensures greater accountability of the data controllers.

5.7.2 Personal Data Stores

The European Commission Communication on Big Data has suggested and encouraged the use of personal data spaces or personal data stores which complements data portability and allows the data subject to control who has access to his/her data and for what purpose it is being used.[164] These data spaces are commonly used to store real-time data such as location tracked by GPS and blood pressure which can be measured by a fitness tracker.

5.8 Co-Regulation

My Data Approach

The Nordic countries have adopted a new approach called "My Data" which seeks to shift the control over personal data and processing from the organizations to the data subjects.[165] The aim of this approach is to provide the data subjects with greater access to their personal information in a machine-readable and open format, which will help them manage and control the use of their data in a better and streamlined way.[166]

5.9 Self-Regulation

5.9.1 Agonistic pluralism of algorithms

Crawford (2016) suggests agonistic pluralism as both a design ideal for engineers, as well as a provocation to understanding algorithms in a broader social context in order to challenge the algorithmic "black box". According to Crawford, rather than focusing on the calculations in isolation, we need to account for the spaces of contestation in which they operate.[167] In isolation, many of these algorithms seem the opposite of agonistic: much of the complexity of search, ranking and recommendation algorithms is non-negotiable and kept far from view, inside an algorithmic "black box". If algorithms adopt deliberative democratic paradigms, they assume an internal mechanism of equal agents, rational debate and emerging consensus positions.

5.9.2 Institutional Review Boards (IRBs)

IRBs are review bodies that govern research, which includes human subjects and sensitive datasets. With collection of personal data on a real-time basis in the age of big data, IRBs play a significant role in self-regulating bodies that collect and process information in order to avoid unfair and unwarranted use of such personal data, which might be detrimental to the data subjects.[168]

5.9.3 Code of Conduct

It has been suggested that the Mobile Network Operators adopt a Code of Conduct which involves embedding standard procedures with safeguards such as anonymization or pseudonymization techniques. This would prevent big data harms and minimize the costs of transaction while transmitting the data to third parties who will then use the data for public purposes.[169]

5.9.4 Responsible Data

A few companies have voluntarily taken up practices which safeguard personal information from voluminous amounts of data in order to ensure that the data being processed and analysed does not result in any harm to the data subject. The Responsible Data Forum focuses on tactics and methodologies to extract the benefits from big data analytics while ensuring fair and reasonable use of personal information of the data subject.[170]

6. Deregulation

6.1 Open Data Policy

Various states and countries have policies which promote open data, but these jurisdictions do not have policies which seek to deregulate existing laws in order to benefit from the use of big data analytics in particular.

6.1.1 Shared Data

Shared data facilitates the development of accountability systems by making available anonymized microdata to Civil Society Organisations, Academics, and other stakeholders.[171] The concept of "Semi-Open" Data helps researchers and scholars gain access to sensitive data, which serves their purposes, while maintaining anonymity with other parties. This is done on the condition that the researchers do not share such sensitive information. Data can also be released on a large scale on an agreement or license that requires the re-users to refrain from re-identifying the data.

6.1.2 Metatransparency

This is the practice of incorporating transparency into governance structures of open data programs.[172] Different kinds of personal information are susceptible to different degrees of privacy risks. Metadata helps identify the nature, source and the privacy risks that come along with different kinds of data.

7. Ethical standards

There is a need to develop data ethics standards to complement regulation and mitigate big data harms. This will require sector-specific expertise as well as a grassroots perspective so that data scientists are trained to identify and minimise biases in datasets. Definitions of fairness, anti-discrimination, and bias in society have been debated by social scientists extensively. The computer science community has recently been engaged in a similar debate, attempting to quantify different decisions-making philosophies taking into account disparate impact and disparate treatment.173 Michael Feldman and others have attempted to develop a mathematical criteria based on the US Equal Employment Opportunity Commission (EEOC)'s 80 percent rule, in order to identify disparate impact174 in a dataset and removing it.175 Literature around regulatory frameworks for ethics and big data have focussed on the following:

1. **Bringing data science under ethical regulation:** Metcalf and Crawford have noted discontinuities between data science and established tools of research ethics regulation.[176] As the disciplines that have so far informed data science, i.e., computer science, physics, and applied mathematics, have been about systems and not human beings, therefore they have been outside the scope of human-related-ethics concerns and regulations. However, there is a growing realization that data science has a very clear impact on the lives of people, and is not equipped to address ethical challenges.
2. **Ethical guidelines for data research:** Urs Gasser et al have pointed the growing trend of big data research being carried out beyond traditional oversight mechanisms.[177] They propose developing a new ethical framework in a multi-stakeholder fashion that would not only be guided by human rights but also the right of individuals to participate in the production of scientific knowledge. According to Zimmer, conceptual gaps in big data research, ethical dilemmas inherent in research projects, outreach efforts and developing policy guidance must be looked at by the ethical, research and regulatory communities. He argues that the ethical dimensions of big data can only be understood and addressed once these issues are studied.[178] Tene and Polonetsky suggest that the threshold to determine whether or not research should be subject to ethical principles should be to see whether it is research affecting individuals, and if it is, then it must be subject to ethical principles regardless of whether or not personally identifiable information is used.[179]
3. **Ethics norms:** Shilton argues that researchers will never be entirely self regulating, and ethical norms that currently exist among researchers go beyond formal requirements of IRBs, including increasing transparency and engaging in deliberative ethics processes.[180] According to Urs Gasser et al, the multi-stakeholder group that would develop an ethical framework for big data (discussed above) should consider developing ethical norms based partly on existing best practices for research ethics.[181]
4. **Ethics and trust:** Richards and Hartzog advocate that for organizations to truly protect data subjects, they must embrace the notion of trust as a guiding principle in their processing of personal data, and structure procedures around affirmative obligations to ensure they act as adequate stewards of the data with which they are entrusted.[182]

8. Conclusion

Today big data presents us with as many challenges as it does benefits. Whilst big data analytics can offer incredible opportunities to reduce inefficiency, improve decision-making and increase transparency, concerns remain about the effects of these new technologies on issues such as privacy, equality and discrimination. Although the tensions between the competing demands of big data advocates and their critics may appear irreconcilable; only by highlighting these points of contestation can we hope to begin to ask the types of important and difficult questions necessary to do so. These include: how can we reconcile big data's need for massive inputs of personal information with core principles of privacy such as data minimization and collection limitation? What processes and procedures need to be instilled during the design and implementation of big data models and algorithms to provide sufficient transparency and accountability so as to avoid instances of discrimination? What measures can be used to help bridge digital divides and ensure that the benefits of big data are shared equitably? Questions such as these are today only just beginning to be addressed. Each one however, will require careful consideration and reasoned debate if big data is to deliver on its promises and truly fulfil its 'revolutionary' potential.

ENDNOTES

1 Gantz, J. and Reinsel, D. "Extracting Value from Chaos". IDC. June 2011. Accessed on December 1, 2016. http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf

2 Meeker, Mary and Wu, Liang. "Internet Trends D11 Conference". Kleiner Perkins Caulfield Byers. May 29, 2013. Accessed December 1, 2016.http://www.slideshare.net/kleinerperkins/kpcb-internet-trends-2013.

3 Doug, Laney. "3D Data Management: Controlling Data Volume, Velocity and Variety". Gartner. February 6, 2001. Accessed on December 2, 2016 https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

4 Boyd, Danah, and Crawford, Kate "Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon". Information, Communication & Society,Vol 15, Issue 5. Routledge. May 10, 2012. Accessed on December 2, 2016. http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878, Tene, O., & Polonetsky, J. "Big Data for All: Privacy and User Control in the Age of Analytics". *11 Nw. J. Tech. & Intell. Prop. 239* (2013). Accessed on December 2, 2016. http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

5 Boyd, Danah, and Crawford, Kate "Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon". Information, Communication & Society,Vol 15, Issue 5. Routledge. May 10, 2012. Accessed on December 2, 2016. http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878, Tene, O., & Polonetsky, J. "Big Data for All: Privacy and User Control in the Age of Analytics". *11 Nw. J. Tech. & Intell. Prop. 239* (2013). Accessed on December 2, 2016. http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

6 McKinsey Global institute. “Big data: The next frontier for innovation, competition, and productivity”. June 2011. McKinsey. Accessed on December 2, 2016. http://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx

7 Joh. Elizabeth E. “Policing by Numbers: Big Data and the Fourth Amendment”. Washington Law Review, Vol. 85: 35. 2014. Accessed on December 2, 2016. https://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1319/89WLR0035.pdf?sequence=1

8 Raghupathi, Wullianallur and Raghupathi, Viju. “Big data analytics in healthcare: promise and potential”. BioMed Central. Health Information Science and Systems. January 5, 2014. Accessed on December 2, 2016. https://hissjournal.biomedcentral.com/articles/10.1186/2047-2501-2-3

9 Manyika, James; Chui, Michael; Brown, Brad; Bughin, Jacques; Dobbs, Richard; Roxburgh, Charles; Hung, Angela and Byers. “Big data: The next frontier for innovation, competition, and productivity”. May 2011. Accessed on December 2, 2016. http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

10 Manyika, James; Chui, Michael; Brown, Brad; Bughin, Jacques; Dobbs, Richard; Roxburgh, Charles; Hung, Angela and Byers. “Big data: The next frontier for innovation, competition, and productivity”. May 2011. Accessed on December 2, 2016. http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

11 Anderson, Richard and Roberts, Daniel. “Big Data: Strategic Risks and Opportunities”. Crowe Horwarth Global Risk Consulting Limited. September 2012. Accessed on December 2, 2016. https://www.crowehorwath.net/uploadedfiles/crowe-horwath-global/tabbed_content/big%20data%20strategic%20risks%20and%20opportunities%20white%20paper_risk13905.pdf

12 Anderson, Richard & Roberts, Daniel. “Big Data: Strategic Risks and Opportunities”. Crowe Horwarth Global Risk Consulting Limited. September 2012. Accessed on December 2, 2016. https://www.crowehorwath.net/uploadedfiles/crowe-horwath-global/tabbed_content/big%20data%20strategic%20risks%20and%20opportunities%20white%20paper_risk13905.pdf

13 Kshetri. N. “The Emerging role of Big Data in Key development issues: Opportunities, challenges, and concerns”. Big Data & Society, Sage. 2014. Accessed on December 2, 2016. http://bds.sagepub.com/content/1/2/2053951714564227.abstract

14 Tene, Omer & Polonetsky, Jules. “Big Data for All: Privacy and User Control in the Age of Analytics”. *11 Nw. J. Tech. & Intell. Prop. 239*. 2013. Accessed on December 2, 2016. http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

15 “IoE-Driven Smart Street Lighting Project Allows Oslo to Reduce Costs, Save Energy, Provide Better Service”. Cisco. 2014. Accessed on December 2, 2016. http://www.cisco.com/c/dam/m/en_us/ioe/public_sector/pdfs/jurisdictions/Oslo_Jurisdiction_Profile_051214REV.pdf

16 Newell. B. C. “Local Law Enforcement Jumps on the Big Data Bandwagon: Automated License Plate Recognition Systems, Information Privacy, and Access to Government Information”. University of Washington - the Information School. October 16, 2013. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2341182

17 Manyika, James; Chui, Michael; Brown, Brad; Bughin, Jacques; Dobbs, Richard; Roxburgh, Charles; Hung, Angela and Byers. “Big data: The next frontier for innovation, competition, and productivity”. May 2011. Accessed on December 2, 2016. http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

18 Morris, David. “Big data could improve supply chain efficiency-if companies would let it”. Fortune. August 5 2015. Accessed on December 3, 2016. http://fortune.com/2015/08/05/big-data-supply-chain/

19 Manyika, James; Chui, Michael; Brown, Brad; Bughin, Jacques; Dobbs, Richard; Roxburgh, Charles; and Byers, Angela Hung. “Big Data: The next frontier for innovation, competition, and productivity”. May 2011. Accessed on December 3, 2016. http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

20 Tucker, Darren S., & Wellford, Hill B. “Big Mistakes Regarding Big Data”. Antitrust Source, American Bar Association. 2014. SSRN: http://ssrn.com/abstract=2549044

21 Davenport, T., Barth., Bean, R. “How is Big Data Different”. MITSloan Management Review, Fall 2012. http://sloanreview.mit.edu/article/how-big-data-is-different/

22 Manyika, James; Chui, Michael; Brown, Brad; Bughin, Jacques; Dobbs, Richard; Roxburgh, Charles; and Byers, Angela Hung. “Big Data: The next frontier for innovation, competition, and productivity”. May 2011. Accessed on December 3, 2016. http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

23 “Big Data, Digital Media, and Computational Social Science: Possibilities and Perils”. Shah, Dhavan V., Cappella, Joseph N., and Neum, W. Russell. University of Wisconsin-Madison. May 2015. https://www.journalism.wisc.edu/~dshah/ANNALS2015a.pdf

24 Tucker, Darren S., & Wellford, Hill B.,“Big Mistakes Regarding Big Data”. Antitrust Source, American Bar Association. 2014. SSRN. Accessed on December 3, 2016. http://ssrn.com/abstract=2549044

25 Raghupathi, Wullianallur and Raghupathi, Viju. “Big data analytics in healthcare: promise and potential”. BioMed Central. *Health Information Science and Systems*. January 5, 2014. Accessed on December 2, 2016. https://hissjournal.biomedcentral.com/articles/10.1186/2047-2501-2-3

26 Brown, B., Chui, M., Manyika, J. “Are you Ready for the Era of Big Data?”. McKinsey Quarterly. 2011. http://www.t-systems.com/solutions/download-mckinsey-quarterly-/1148544_1/blobBinary/Study-McKinsey-Big-data.pdf ; Benady, D. “Radical transparency will be unlocked by technology and big data”. Guardian 2014. http://www.theguardian.com/sustainable-business/radical-transparency-unlocked-technology-big-data

27 Brown, B., Chui, M., Manyika, J. ‘Are you Ready for the Era of Big Data?’, McKinsey Quarterly, (2011), Available at http://www.t-systems.com/solutions/download-mckinsey-quarterly-/1148544_1/blobBinary/Study-McKinsey-Big-data.pdf ; Benady, D., ‘Radical transparency will be unlocked by technology and big data’, Guardian (2014) Available at: http://www.theguardian.com/sustainable-business/radical-transparency-unlocked-technology-big-data

28 Brown, B., Chui, M., Manyika, J. “Are you Ready for the Era of Big Data?”. McKinsey Quarterly. 2011. http://www.t-systems.com/solutions/download-mckinsey-quarterly-/1148544_1/blobBinary/Study-McKinsey-Big-data.pdf; Benady, D., ‘Radical transparency will be unlocked by technology and big data’. Guardian. 2014. http://www.theguardian.com/sustainable-business/radical-transparency-unlocked-technology-big-data

29 Independent Expert Advisory Group on a Data Revolution for Sustainable Development. United Nations, "A World That Counts: Mobilising the Data Revolution for Sustainable Development, Report prepared at the request of the United Nations Secretary-General". November 2014. Accessed on December 3, 2016. http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf,

30 Independent Expert Advisory Group on a Data Revolution for Sustainable Development. United Nations, "A World That Counts: Mobilising the Data Revolution for Sustainable Development, Report prepared at the request of the United Nations Secretary-General". November 2014. Page 18. Accessed on December 3, 2016. http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf. See also, Hilbert, M. "Big Data for Development: From Information- to Knowledge Societies". 2013. SSRN. http://ssrn.com/abstract=2205145

31 Richards, Neil M. & King, Jonathan H. "Three Paradoxes of Big Data". September 2013. https://www.stanfordlawreview.org/online/privacy-and-big-data-three-paradoxes-of-big-data/

32 Tene, O., & Polonetsky, J. "Big Data for All: Privacy and User Control in the Age of Analytics". 11 Nw. J. Tech. & Intell. Prop. 239. 2013. http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

33 Narayanan, Arvind & Shmatikov, Vitaly. "Robust De-anonymization of Large Sparse Datasets". 2008. IEEE SYMP. ON SECURITY & PRIVACY 111. Quoted in Tene, O., & Polonetsky, J. "Big Data for All: Privacy and User Control in the Age of Analytics". 11 Nw. J. Tech. & Intell. Prop. 239 2013 http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

34 Government of India, Planning Commission, "Report of the Group of Experts on Privacy" (Chaired by Justice A P Shah, Former Chief Justice, Delhi High Court) , October 2012. http://planningcommission.nic.in/reports/genrep/rep_privacy.pdf

35 OECD. "Guidelines on the Protection of Privacy and Transborder Flows of Personal Data". The Organization for Economic Co-Operation and Development. 1999.; The European Parliament and the Council of the European Union, "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data" EU Data Protection Directive. 1995

36 Barocas, S., & Selbst, A.D., "Big Data's Disparate Impact".California Law Review. *Vol. 104*. 2015. Accessed on December 3, 2016. http://ssrn.com/abstract=2477899

37 Article 29 Working Group. "Opinion 03/2013 on purpose limitation". Article 29 Data Protection Working Party. 2 April 2013. Accessed on December 3, 2016. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

38 Solove, Daniel J. "Privacy Self-Management and the Consent Dilemma". 126 Harv. L. Rev. 1880. 2013. Accessed on December 3, 2016. http://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2093&context=faculty_publications

39 Brandimarte, L., Acquisti, A., & Loewenstein, G., "Misplaced Confidences: Privacy and the Control Paradox". Ninth Annual Workshop on the Economics of Information Security (WEIS). June 7-8, 2010, Harvard University, Cambridge, MA. Accessed on December 3, 2016. https://fpf.org/wp-content/uploads/2010/07/Misplaced-Confidences-acquisti-FPF.pdf

40 Solove, D, J., "Privacy Self-Management and the Consent Dilemma". 126 Harv. L. Rev. 1880 (2013). Accessed on December 3, 2016. http://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2093&context=faculty_publications

41 Yu, W, E., Data. "Privacy and Big Data-Compliance Issues and Considerations". ISACA Journal, Vol. 3 2014 http://www.isaca.org/Journal/archives/2014/Volume-3/Pages/Data-Privacy-and-Big-Data-Compliance-Issues-and-Considerations.aspx

42 Ramirez, E., Brill, J., Ohlhausen, M., Wright, J., & McSweeny, T. "Data Brokers: A Call for Transparency and Accountability". Federal Trade Commission (2014). Accessed on December 3, 2016. https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

43 Huber, Rohan. "Mosaic Effect Paints Vivid Pictures of Tech Users Lives, Felten Tells Privacy Board." Woodrow Wilson School of Public & International Affairs Princeton University. November 20th 2014. Available at: http://wws.princeton.edu/news-and-events/news/item/mosaic-effect-paints-vivid-pictures-tech-users-liairdsves-felten-tells-privacy

44 Foucault, Michel. "Discipline and Punish: The Birth of the Prison". Translated by Alan Sheridan, London: Allen Lane, Penguin (1977).

45 Marthews, A., & Tucker, C., "Government Surveillance and Internet Search Behavior". (2015) SSRN. Accessed on December 3, 2016. http://ssrn.com/abstract=2412564

46 Boyd, D., and Crawford, K. "Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon". Information, Communication & Society, Vol. 15, Issue 5, (2012).

47 Hirsch, D., "That's Unfair! Or is it? Big Data, Discrimination and the FTC's Unfairness Authority". Kentucky Law Journal, Vol. 103. http://www.kentuckylawjournal.org/wp-content/uploads/2015/02/103KyLJ345.pdf

48 Hill, K., "How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did". February 16, 2012. Accessed on December 3, 2016.

http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

49 Barocas, Solon and Selbst, Andrew D. "Big Data's Disparate Impact". California Law Review 2016. Accessed on December 3, 2016.

http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf

50 Robinson, D., Yu, H. "Civil Rights, Big Data, and Our Algorithmic Future". September 2014. http://bigdata.fairness.io/introduction/

51 Schrange, Michael. "Big Data's Dangerous New Era of Discrimination". Harvard Business Review.

52 Robinson, D., Yu, H. "Civil Rights, Big Data, and Our Algorithmic Future". September 2014. http://bigdata.fairness.io/introduction/

53 Robinson, D., Yu, H. "Civil Rights, Big Data, and Our Algorithmic Future". September 2014. http://bigdata.fairness.io/introduction/

54 Schradie, J. “Big Data Not Big Enough? How the Digital Divide Leaves People Out”. MediaShift. July 31, 2013. http://mediashift.org/2013/07/big-data-not-big-enough-how-digital-divide-leaves-people-out/

55 Tara Thigarajan, India's challenge of 'dark data', Livemint, available at http://www.livemint.com/Opinion/JZ95vbh4jS5KI1ll4kjcIO/Indias-challenge-of-dark-data.html.

56 Crawford, K. “The Hidden Biases in Big Data”. Harvard Business Review. April 1, 2013. Accessed on December 3, 2016.https://hbr.org/2013/04/the-hidden-biases-in-big-data

57 Boyd, D., and Crawford, K. “Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon”. Information, Communication & Society.Vol 15, Issue 5. March 20, 2012. Accessed on December 3, 2016. http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878; Newman, N., “Search, Antitrust and the Economics of the Control of User Data”. 31 YALE J. ON REG. 401 (2014).

58 Jonas, Lerman. “Big Data and its Exclusions”. September 2013. https://www.stanfordlawreview.org/online/privacy-and-big-data-big-data-and-its-exclusions/

59 Yadav, Anumeha. “In Rajasthan, there is 'unrest at the ration shop' because of error-ridden Aadhaar”. Scroll.in. April 2, 2016.

http://scroll.in/article/805909/in-rajasthan-there-is-unrest-at-the-ration-shop-because-of-error-ridden-aadhaar

60 Chattapadhyay, Sumandro. “The Last Chance for a Welfare State Doesn't Rest in the Aadhaar System”. The Wire. April 19, 2016. Accessed on December 3, 2016. http://thewire.in/30256/the-last-chance-for-a-welfare-state-doesnt-rest-in-the-aadhaar-system/

61 Price, Megan; Ball, Patrick. “Big Data, Selection Bias, and the Statistical Patterns of Mortality in Conflict”.

62 Barocas, S., & Selbst, A, D., “Big Data's Disparate Impact”.California Law Review, Vol. 104. (2015). SSRN: http://ssrn.com/abstract=2477899; Kshetri. N, “The Emerging role of Big Data in Key development issues: Opportunities, challenges, and concerns”. *Big Data & Society* (2014) http://bds.sagepub.com/content/1/2/2053951714564227.abstract

63 Pasquale, F. “The Black Box Society: The Secret Algorithms That Control Money and Information”. Harvard University Press. (2015).

64 Shapiro, Carl and Varian, Hal R. “Information Rules”. Havard Business School Press. 1999. Accessed on December 4, 2016.

http://www.uib.cat/depart/deeweb/pdi/acm/arxius/premsa/information-rules%20VARIAN%20SHAPIRO.pdf

65 Newman, N. “The Cost of Lost Privacy: Search, Antitrust and the Economics of the Control of User Data”. 2013. SSRN. http://ssrn.com/abstract=2265026, Newman, N. “Search, Antitrust and the Economics of the Control of User Data”. 31 YALE J. ON REG. 401. 2014.

66 Newman, N. “The Cost of Lost Privacy: Search, Antitrust and the Economics of the Control of User Data”. 2013. SSRN. http://ssrn.com/abstract=2265026, Newman, N. “Search, Antitrust and the Economics of the Control of User Data”. 31 YALE J. ON REG. 401. 2014.

67 European Data Protection Supervisor, “Privacy and competitiveness in the age of big data:The interplay between data protection, competition law and consumer protection in the Digital Economy”, March 2014. https://secure.edps.europa.eu/EDPSWEB/webdav/shared/Documents/Consultation/Opinions/2014/14-03-26_competitition_law_big_data_EN.pdf.

68 Inukollu, Venkata. Security Issues Associated with Big Data in Cloud Computing. International Journal of Network Security & Its Applications, Vol 6, No 3. May 2014. Available at: https://pdfs.semanticscholar.org/3e0a/548fae69482e550c21a6dc7806213203c3bc.pdf

69 Ferguson, Mike. Enterprise Information Protection - The Impact of Big Data. Intelligent Business Strategies. Available at: http://www-01.ibm.com/software/os/systemz/pdf/Info_Security_and_Big_Data_on_Z_White_Paper_Final.pdf

70 Anderson, C. “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete”. WIRED. June 23 2008. Accessed on December 4, 2016. www.wired.com/2008/06/pb-theory/

71 Cukier Kenneth., Mayer-Schoenberger, Viktor. The Rise of Big Data: How it's changing the way we think about the world. Foreign Affairs 92. May / June 2013. Accessed on December 4, 2016. https://www.foreignaffairs.com/articles/2013-04-03/rise-big-data

72 Boyd, D. and Crawford, K. “Critical questions for big data. Information, Communication and Society” 15(5): 662-679. Routledge. May 2012. Accessed on December 4, 2016. https://people.cs.kuleuven.be/~bettina.berendt/teaching/ViennaDH15/boyd_crawford_2012.pdf; Kitchen, R. (2014) Big Data, new epistemologies and paradigm shifts, Big Data & Society. Sage. April-June 2014: 1-12. http://progcity.maynoothuniversity.ie/2014/06/new-paper-big-data-new-epistemologies-and-paradigm-shifts/

73 Frické, M. “Big Data and its epistemology”. Journal of the Association for Information Science and Technology 66(4): 651–661.The University of Arizona. 2013. Accessed on December 4, 2016. https://ischool.arizona.edu/sites/si.arizona.edu/files/FrickeBigDataPaperShorterFormat.pdf

74 Kitchen, R (2014) Big Data, new epistemologies and paradigm shifts, Big Data & Society, April-June 2014: 1-12

75 King quoted in Shaw, J. “Why Big Data is a Big Deal”. Harvard Magazine March-April 2014. http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal

76 Gross quoted in Kitchen, R. “Big Data, new epistemologies and paradigm shifts, Big Data & Society”. April-June 2014: 1-12.

77 Hey, T., Tansley, S., and Tolle, K (eds). “The Fourth Paradigm: Data-Intensive Scientific Discovery”. Redmond: Microsoft Research, pp. xvii-xxxi.

78 Hey, T., Tansley, S., and Tolle, K. (eds). “The Fourth Paradigm: Data-Intensive Scientific Discovery”. Redmond: Microsoft Research, pp. Xvii-xxxi. 2009.

79 Mayer-Schonberger, V and Cukier K. “Big Data: A Revolution that Will Change How We Live, Work and Think”. London: John Murray.

80 Joh. E. “Policing by Numbers: Big Data and the Fourth Amendment”. Washington Law Review, Vol. 85: 35. 2014. https://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1319/89WLR0035.pdf?sequence=1

81 Bowker, G. “The Theory-Data Thing”. International Journal of Communication 8 (2043), 1795-1799.

82 Boyd, D. and Crawford, K. “Critical questions for big data”. Information, Communication and Society 15(5): 662-679.

83 Boyd, D. and Crawford, K. "Critical questions for big data". Information, Communication and Society 15(5): 662-679. 2012; Kitchen, R. "Big Data, new epistemologies and paradigm shifts". Big Data & Society. April-June 2014: 1-12.

84 McCue, C., "Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis". Butterworth-Heinemann. 2014.

85 Dyche, J. "Big data 'Eurekas!' don't just happen". Harvard Business Review Blog. November 20. http://blogs.hbr.org/cs/2012/11/eureka_doesnt_just_ happen.html

86 Steadman I "Big data and the death of the theorist". Wired, 25 January 2013. http://www.wired.co.uk/news/archive/2013-01/25/big-data-end-of-theory

87 Prensky, M. "H. sapiens digital: From digital immigrants and digital natives to digital wisdom". Innovate 5(3), http://www.innovateonline.info/index.php?view¼article&id¼705

88 Boyd, D., and Crawford, K. "Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon", Information, Communication & Society,Vol 15, Issue 5.2012. http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878

89 Boyd, D. and Crawford, K. "Critical questions for big data. Information, Communication and Society". 2012 15(5): 662-679.

90 Griffin, James. "On Human Rights". January 1, 2008, Oxford.

91 Public Opinion, Walter Lippmann

92 The Phantom Public, Walter Lippmann

93 Sandvig et al., 2014.; Seaver, 2014.

94 Burrell, Jenna. "How the machine 'thinks': Understanding opacity in machine learning algorithms". 2016. http://bds.sagepub.com/content/spbds/3/1/2053951715622512.full.pdf

95 Ziewitz, Malte. "Evaluation as Governance: The Practical Politics of Reviewing, Rating and Ranking on the Web." DPhil Thesis. University of Oxford. 2012. http://zwtz.org/projects/evaluation-as-governance.html

96 Barocas, Solon; Hood, Sophie and Ziewitz, Malte. "Governing Algorithms: A Provocation Piece". http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2245322.

97 Schadt, Dr. Eric and Chilukuri, Sastry. "The role of big data in medicine". http://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/the-role-of-big-data-in-medicine.

98 Pomerleau, Mark. "6 big data projects to aid disaster response". GCN, May 4, 2015. https://gcn.com/articles/2015/05/04/big-data-disaster-nsf.aspx.

99 Turner, M.; Walker, P.; Chaudhuri, S. and Varghese, R. "A New Pathway to Financial Inclusion: Alternative Data, Credit Building, and Responsible Lending in the Wake of the Great Recession." Policy and Economic Research Council. 2012. http://www.perc.net/wp-content/uploads/2013/09/WEB-file-ADI5-layout1.pdf.

100 The U.S. Department of Justice's Office of Community Oriented Policing Services. "Final Report of the President's Task Force on 21st Century Policing". May 2015. http://www.cops.usdoj.gov/pdf/taskforce/taskforce_finalreport.pdf.

101 Stone, Martha. "Big Data for Media". Reuters Institute for the study of Journalism. November 2014. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Big%20Data%20For%20Media_0.pdf.

102 Pittman, David. "Big Data in Retail: Examples in Action". IBM. Accessed on December 4, 2016. http://www.ibmbigdatahub.com/presentation/big-data-retail-examples-action.

103 Ungerleider, Neal. "Colleges are using Big Data to predict which students will do well before they accept them". Fast Company. October 21, 2013. Accessed on December 4, 2016. http://www.fastcoexist.com/3019859/futurist-forum/colleges-are-using-big-data-to-predict-which-students-will-do-well-before-the.

104 Sassen, Saskia. "Talking back to your intelligent city". Mckinsey on Society. Accessed on December 4, 2016. http://voices.mckinseyonsociety.com/talking-back-to-your-intelligent-city/.

105 Saskia Sassen, "The Future of Smart Cities". Accessed on December 4, 2016. http://videos.liftconference.com/video/2895375/saskia-sassen-the-future-of-smart-cities.

106 Burrell, Jenna. "How the machine 'thinks': Understanding opacity in machine learning algorithms". Sage. 2016. http://bds.sagepub.com/content/spbds/3/1/2053951715622512.full.pdf

107 Pedro Domingos, The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World, Basic Books, 2015 at 376.

108 Maryanne Gaitho, How Applications of Big Data Drive Industries, available at https://www.simplilearn.com/big-data-applications-in-industries-article.

109 Quentin Baltus, Understanding big data key debates and controversies: a critical literature review, available at https://www.linkedin.com/pulse/understanding-big-data-key-debates-controversies-critical-baltus

110 Kobie, Nicole. "Who do you blame when an algorithm gets you fired?" WIRED. January 29, 2016. Accessed on December 4, 2016. http://www.wired.co.uk/article/make-algorithms-accountable

111 Feldman, Michael et al., "Certifying and Removing Disparate Impact". July 16, 2015. http://arxiv.org/abs/1412.3756

112 Crawford, Kate & Schultz, Jason. "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms". Boston College Law Review. http://lawdigitalcommons.bc.edu/cgi/viewcontent.cgi?article=3351&context=bclr

113 Crawford, Kate & Schultz, Jason. "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms". Boston College Law Review. http://lawdigitalcommons.bc.edu/cgi/viewcontent.cgi?article=3351&context=bclr

114 Article 22, GDPR

115 Danielle K. Citron Technological Due Process, 85 WASH. U. L. REV. 1249, 1256 (2008)

116 Danielle K. Citron, "Technological Due Process". 85 WASH. U. L. REV. 1249, 1256 (2008)

117 "Privacy by design in big data". ENISA. December 17, 2015. Accessed on December 4, 2016. https://www.enisa.europa.eu/publications/big-data-protection

118 Pearson, Siani and Mont, Marco Cassasa. "HP Labs Bristol". IEEE Computer Society https://documents.epfl.ch/users/a/ay/ayday/www/mini_project/Sticky%20Policies.pdf

119 Pearson, Siani and Mont, Marco Cassasa. "HP Labs Bristol". IEEE Computer Society. https://documents.epfl.ch/users/a/ay/ayday/www/mini_project/Sticky%20Policies.pdf

120 Privacy by design in big data: ENISA (Report), https://www.enisa.europa.eu/publications/big-data-protection

121 Privacy by design in big data: ENISA (Report), https://www.enisa.europa.eu/publications/big-data-protection

122 Australian Privacy Amendment (Re-identification offence) Bill, 2016

123 GDPR Article 22

124 GDPR Article 19 (ICO), https://ico.org.uk/media/1541/big-data-and-data-protection.pdf

125 Article 20 (ICO). "Big data and data protection". Information Commissioner's Office. Accessed on December 5, 2016. https://ico.org.uk/media/1541/big-data-and-data-protection.pdf

126 "Big data and data protection". Information Commissioner's Office. Accessed on December 5, 2016. https://ico.org.uk/media/1541/big-data-and-data-protection.pdf

127 Fred Cate, Government Data Mining: The Need for a Legal Framework, available at http://www.law.harvard.edu/students/orgs/crcl/vol43_2/435-490_Cate.pdf.

128 "Big data and data protection". Information Commissioner's Office. Accessed on December 5, 2016. https://ico.org.uk/media/1541/big-data-and-data-protection.pdf

129 Executive Office of the President. "Big Data: Seizing Opportunities, Preserving Values". The White House. Accessed on December 5, 2016. https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

130 Article 4 (3b), GDPR

131 Daniel Solove, Privacy self-management and consent dilemma, 2013 available at http://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2093&context=faculty_publications.

132 Article 5(c). GDPR. "Big data and data protection". Information Commissioner's Office. Accessed on December 5, 2016. https://ico.org.uk/media/1541/big-data-and-data-protection.pdf

133 Article 29. "Data Protection Working Party, Opinion 03/2013 on purpose limitation". 2013. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

134 Article 14. GDPR. "Big data and data protection". Information Commissioner's Office. Accessed on December 5, 2016. https://ico.org.uk/media/1541/big-data-and-data-protection.pdf

135 "Big data and data protection". Information Commissioner's Office. Accessed on December 5, 2016. https://ico.org.uk/media/1541/big-data-and-data-protection.pdf

136 Big Data: Seizing Opportunities, Preserving Values, White House , https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

137 The Federal Trade Commission. "Big Data: A Tool for Inclusion or Exclusion?" January 2016. https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf

138 Barocas, Solon and Selbst, Andrew D. "Big Data's Disparate Impact". California Law Review. 2016. http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf

139 Hoffman, Sharona. "Big Data and the Americans with Disabilities Act". Faculty Publications. 2017. http://scholarlycommons.law.case.edu/cgi/viewcontent.cgi?article=2986&context=faculty_publications

140 Genetic Information Nondiscrimination Act

141 Burdon, Mark and Harpur, Paul. "Re-conceptualising Privacy and Discrimination in an age of talent analytics". UNSW Law Journal, Volume 37(2), 679-712. http://www.unswlawjournal.unsw.edu.au/sites/default/files/t2_burdon_and_harpur.pdf

142 Ohlhausen, Maureen K. and Okuliar, Alexander P. "Competition, Consumer Protection and the Right (Approach) to Privacy". Antitrust Law Journal. 2015. Accessed on December 5, 2016. https://www.ftc.gov/system/files/documents/public_statements/686541/ohlhausenokuliaralj.pdf

143 *See generally* Pamela Jones Harbour, Comm'r, Fed. Trade Comm'n, Dissenting Statement, Google/DoubleClick, FTC File No. 071-0170, at 1 (2007). *available at* www.ftc.gov/sites/default /files/documents/public_statements/statement-matter-google/doubleclick/071220harbour_0.pdf.

144 Google/DoubleClick Merger case

145 French Competition Authority, Opinion n°10-A-13 of 1406.2010, http://www.autoritedelaconcurrence.fr/pdf/avis/10a13.pdf. That opinion of the Authority aimed at giving general guidance on that subject. It did not focus on any particular market or industry although it described a possible application of its analysis to the telecom industry.

146 Brill, Julie Comm'r, Fed. Trade Comm'n, *The Intersection of Consumer Protection and Competition in the New World of Privacy*, COMPETITION POL'Y INT'L, Spring 2011, at 7, 10. Commissioner Brill has also written that the 2010 Merger Guidelines offer federal agencies "am- ple room to consider the impact of a transaction on privacy-based competition." Julie Brill, *Com- petition and Consumer Protection: Strange Bedfellows or Best Friends?*, ANTITRUST SOURCE, Dec. 2010, at 10, www.abanet.org/antitrust/at-source/10/12/Dec10-Brill12-21f.pdf.

147 *See generally* J. Thomas Rosch, Concurring and Dissenting Statement of Commissioner J. Thomas Rosch Regarding Google's Search Practices, Google Inc., FTC File No. 111-0163 (Jan. 3, 2013), *available at* www.ftc.gov/sites/default/files/documents/public_statements/concurring- and-dissenting-statement-commissioner-j.thomas-rosch-regarding-googles-search-practices/ 130103googlesearchstmt.pdf.

148 Canadian Competition Bureau. http://www.competitionbureau.gc.ca/eic/site/cb-bc.nsf/eng/04092.html

149 See the press release on a workshop on Big Data and Competition, held by the Canadian Competition Bureau http://www.competitionbureau.gc.ca/eic/site/cb-bc.nsf/eng/04030.html#section3_4

150 Gov.UK Verify https://www.gov.uk/government/publications/introducing-govuk-verify/introducing-govuk-verify#certified-companies

151 Analysis Group, http://www.analysisgroup.com/is-big-data-a-true-source-of-market-power/#sthash.5ZHmrD1m.dpuf

152 FTC, “Big Data: A Tool for Inclusion or Exclusion?”. January 2016. Accessed on December 5, 2016. https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf?utm_source=govdelivery

153 FTC, “Big Data: A Tool for Inclusion or Exclusion?”. January 2016. Accessed on December 5, 2016. https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf?utm_source=govdelivery

154 FTC, “Big Data: A Tool for Inclusion or Exclusion?”. January 2016. Accessed on December 5, 2016. https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf?utm_source=govdelivery

155 FTC, “Big Data: A Tool for Inclusion or Exclusion?”. January 2016. Accessed on December 5, 2016. https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf?utm_source=govdelivery

156 Grunes, Allen P. & Stucke, Maurice E. “No Mistake About It: The Important Role of Antitrust in the Era of Big Data”, ANTITRUST SOURCE. April 2015.

157 Petkova, Bilyana and Hacker, Philipp, “Reining in the Big Promise of Big Data: Transparency, Inequality, and New Regulatory Frontiers”. 2016. Lecturer and Other A liate Scholarship Series. Paper 13. http://digitalcommons.law.yale.edu/ylas/13

158 “Big Data: Seizing Opportunities, Preserving Values”. White House. May 1, 2014. Accessed on December 5, 2016. https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

159 “Big Data: Seizing Opportunities, Preserving Values”. White House. May 1, 2014. Accessed on December 5, 2016. https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

160 “Big Data: Seizing Opportunities, Preserving Values”. White House. May 1, 2014. Accessed on December 5, 2016. https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

161 European Data Protection Supervisor. “Meeting the challenges of Big Data: A call for transparency, user control, data protection by design and accountability”. Opinion. July 2015.

162 European Data Protection Supervisor. “Meeting the challenges of Big Data: A call for transparency, user control, data protection by design and accountability”. Opinion. July 2015.

163 United Nations Human Rights. “A Human Rights Based Approach to Data”. OHCHR. Accessed December 8, 2016. http://www.ohchr.org/Documents/Issues/HRIndicators/GuidanceNoteonApproachtoData.pdf

164 Commission Communication. “Towards a thriving data-driven economy’, which sets forth the Commission’s strategy on big data”. COM(2014) 442 final.

165 Poikola, Antti; Kuikkaniemi, Kai and Honko, Harri. “MyData – A Nordic Model for human-centered personal data management and processing”. Ministry of Transport & Communications and Open Knowledge Finland. Accessed on December 5, 2016. http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78439/MyData-nordic-model.pdf?sequence=1,

166 Poikola, Antti; Kuikkaniemi, Kai and Honko, Harri. “MyData – A Nordic Model for human-centered personal data management and processing”. Ministry of Transport & Communications and Open Knowledge Finland. Accessed on December 5, 2016. http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78439/MyData-nordic-model.pdf?sequence=1,

167 Crawford, Kate. “Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics”. Science, Technology & Human Values, 41(1), 77-92. April 2016. Accessed on December 5, 2016. http://www.katecrawford.net/docs/CanAnAlgorithmBeAgonistic-April2016.pdf

168 Metcalf, Jacob & Crawford, Kate. “Where are human subjects in Big Data research? The emerging ethics divide”. Sage. June 2016. http://bds.sagepub.com/content/3/1/2053951716650211

169 Samarajiva, Rohan. “LIRNEasia, “Draft Code of Conduct for Third-Party Use of Big Data Generated by Mobile Network Operators”. August 2014. Accessed on December 5, 2016. http://lirneasia.net/wp-content/uploads/2014/08/guidelines-v2.2.pdf

170 Responsible Data, https://www.theengineroom.org/responsible-data/

171 United Nations Human Rights. “A Human Rights Based Approach to Data”. OHCHR. Accessed December 8, 2016. http://www.ohchr.org/Documents/Issues/HRIndicators/GuidanceNoteonApproachtoData.pdf

172 Briefing Paper on Open Data and Privacy. The Centre for Open Data Enterprise. Accessed December 8, 2016. http://reports.opendataenterprise.org/BriefingPaperonOpenDataandPrivacy.pdf

173 Friedler, Sorelle; Scheideger, Carlos and Venkatsubramanian, Suresh. “On the (im)possibility of fairness”. Arxiv. Accessed on November 30, 2016. https://arxiv.org/pdf/1609.07236v1.pdf

174 Barocas, S. and Selbst, A. D. “Big data’s disparate impact”. SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899 Last revised on September 30, 2016. Accessed on December 1, 2016

175 Feldman, Michael; Friedler, Sorelle; Moeller, John; Scheideger, Carlos and Venkatsubramanian, Suresh. “Certifying and removing disparate impact”. July 16, 2015. Arxiv. Accessed on November 30, 2016. https://arxiv.org/pdf/1412.3756v3.pdf

176 Metcalf, Jacob and Crawford, Kate. “Big Data and Society”. Spring. 2016. Accessed November 28, 2016. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2779647

177 Urs Gasser et al. “Towards a New Ethical and Regulatory Framework for Big Data Research”. December 15, 2015. Accessed November 28, 2016.

178 Zimmer, Michael. “Research ethics in the Big Data Era: Addressing Conceptual Gaps for Researchers and IRBs”. December 15, 2015. Accessed December 2, 2016. https://bigdata.fpf.org/papers/research-ethics-in-the-big-data-era-addressing-conceptual-gaps-for-researchers-and-irbs/

179 Tene, Omer and Polonetsky, Jules. “Beyond IRBs: Ethical Guidelines for Data Research.” December 15, 2015. Accessed December 1, 2016. https://bigdatafpf.org/papers/beyond-irbs-ethical-guidelines-for-data-research/

180 Shilton, Katie. “Emerging Ethics Norms in Social Media Research”. December 15, 2015. Accessed December 2, 2016. https://bigdata.fpf.org/papers/emerging-ethics-norms-in-social-media-research/

181 Urs Gasser et al. “Towards a New Ethical And Regulatory Framework for Big Data Research.” December 15, 2015. Accessed November 28,2016. https://bigdata.fpf.org/papers/towards-a-new-ethical-and-regulatory-framework-for-big-data-ethics/

182 Richards, Neil and Hatzgow, Woodrow. “Trusting Big Data Research”. December 15, 2015. Accessed December 2, 2016. https://bigdata.fpf.org/papers/trusting-big-data-research/