religions

MDPI

Article

Big Data, Ethics and Religion: New Questions from a New Science

Michael Fuller

School of Divinity, University of Edinburgh, Mound Place, Edinburgh EH1 2LX, UK; Michael.Fuller@ed.ac.uk;
Tel.: +44-131-650-8963

Academic Editor: Noreen Herzfeld
Received: 30 January 2017; Accepted: 3 May 2017; Published: 10 May 2017

Abstract: Hopes, fears, and ethical concerns relating to technology are as old as technology itself. When considering the increase in the power of computers, and their ever-more widespread use over recent decades, concerns have been raised about the social impact of computers and about practical issues arising from their use: the manner in which data is harvested, the preservation of confidentiality where people's personal information is concerned, the security of systems in which such data is stored, and so on. With the arrival of "big data" new ethical concerns surrounding computer-based technology arise—concerns connected not only with social issues, and with the generation of data and its security, but also with its interpretation by data scientists, and with the burgeoning trade in personal data. The first aim of this paper is to introduce some of these ethical issues, and the second is to suggest some possible ways in which they might be addressed. The latter includes some explorations of the ways in which insights from religious and theological perspectives might be valuable. It is urged that theology and data science might engage in mutually-beneficial dialogue.

Keywords: big data; computing; consent; ethics; hermeneutics; hippocratic oath; interpretation; privacy

1. Introduction

As any technology develops it might be expected that its increasing capabilities give rise to a succession of ethical issues, and computer technology is certainly no exception to this (Tavani 2013a, pp. 6 ff.). Concerns which have been raised in the past include the deskilling of the workforce, increased unemployment, and the health, stress, and isolation of workers, together with issues relating to the storage of personal data in the form of databases (Barbour 1992, pp. 146 ff.). A further concern is the implication of computers in broadening divisions between rich and poor, through the opening up of imbalances between those who have access to computer facilities and the benefits which they bring, and those who do not (Tavani 2013a, p. 305; Barbour 1992, p. 156). All of these concerns are ongoing.

In recent years the increasingly widespread use of computers in all walks of life, from the PCs and smartphones that many consumers use on a daily basis to the supercomputers used in research programmes in astronomy, physics, and medicine, has generated the phenomenon that has become known as "big data":[1] extremely large, often highly heterogeneous, datasets that require novel techniques and new sets of skills to interrogate them. In turn, this has led to a new set of ethical issues surrounding big data. The aim of this paper is to identify, describe, and evaluate some of these ethical

[1] Not least of the issues surrounding big data are linguistic: should these words should be capitalised or not, and should this term should be treated as singular or plural? This paper uses the lower case (except when quoting sources which use the upper), and it follows the convention (given some justification in (Rosenberg 2013, p. 18)) of treating "big data" as a singular form.

issues, and to suggest ways in which some of them may be addressed. It further suggests that insights from the religious domain might be of considerable value in developing these new approaches to big data.

2. Big Data and Data Science

Before going further, it is worth exploring exactly what big data is understood to mean. Although the term is widely used, there is little agreement around a definition, in part because what counts as "big" in this context is changing so rapidly. The description by Laney (2001) in terms of the "three Vs" (volume, variety, and velocity) has been widely quoted: on this understanding, big data is characterised as being concerned with very large quantities of data, which is highly heterogeneous, and which is generated at enormous speed. (Kitchin 2014, p. 68) comments that, in addition to this, big data is exhaustive in scope, fine-grained in resolution, relational in nature (enabling different datasets to be linked), and both flexible and scalable (enabling new fields to be added to it, and the rapid extension of the size of the dataset). Other, broader, definitions have been offered (e.g., (boyd and Crawford 2012, p. 663)).

The distinctiveness of big data goes beyond straightforward issues of size. Mayer-Schönberger and Cukier point out that with the accumulation of so much data, "something new and special is taking place. Not only is the world awash with more information than ever before, but that information is growing faster. The change of scale has led to a change of state. *The quantitative change has led to a qualitative one*" ((Mayer-Schönberger and Cukier 2013, p. 6), my emphasis). Similarly, Kitchin notes that "It is becoming clear that big data have a number of inherent characteristics that make them qualitatively different to previous forms of data" (Kitchin 2014, p. 79). Extremely large datasets are not simply quantitatively different to smaller ones: their sheer scale brings about a "step change", making them qualitatively distinct, too. This means that different tools and different models are required for their handling and analysis.

This, in turn, means that such analysis amounts to a new kind of practice. The development of practices appropriate to the handling of very large datasets has led to the coining of the terms "data science" and "data scientist" (cf. (O'Neil and Schutt 2014, p. 8)) to describe the new techniques which are required, and the practitioners of those techniques. The capabilities required of these practitioners are considerable: according to (Mayer-Schönberger and Cukier 2013, p. 125), the data scientist must combine "the skills of the statistician, software programmer, infographics designer, and storyteller".

These new approaches in turn bring new sets of questions. As boyd and Crawford note, "The age of Big Data has only just begun, but it is important that we start questioning the assumptions, values, and biases of this new wave of research" (boyd and Crawford 2012, p. 675). This is an urgent task, as such assumptions, values and biases may be unhelpful or erroneous, and may have the potential to cause considerable harm.

3. Some Examples of Ethical Concerns Arising from Big Data

There is already a significant literature regarding ethical concerns in information and communication technology (cf. (Tavani 2013b)). What concerns are specific to the big data context? We may immediately note that anyone who engages with services which make use of computers surrenders data to those who run those computers—whether they are consciously aware of it or not. This surrendering of data may occur through engagement with retail or professional services, or through engaging with particular institutions which involve the gathering, analysis, and retention of personal data (e.g., hospitals), or simply through surfing the Internet, or using a mobile phone. A range of issues may then emerge.

3.1. Privacy and Consent

It is standard practice to obtain the informed consent of any party whose data is to be harvested and stored. In practice, consent has generally been obtained through giving information to the individual concerned, usually in the form of a written text, and by that individual signing a form or ticking a box to confirm that they have understood and accept the terms which have been given to them. When using an Internet-based service, such consent is regularly sought by the service providers. In practice, however, it has been observed that this is a process which is geared more towards limiting the liabilities of those harvesting the data rather than genuinely informing the data subjects: "the parties gathering the data typically attempt to minimize the ability of the person about whom the data is being gathered to comprehend the scope of data, and its usage, through a mixture of sharp design and obscure legal jargon" (Wilbanks 2014, p. 235). The practical difficulties of managing privacy and generating informed consent have been summed up as:

> (1) people do not read privacy policies; (2) if people read them, they do not understand them; (3) if people read and understand them, they often lack enough background knowledge to make an informed choice; and (4) if people read them, understand them, and can make an informed choice, their choice may be skewed by various decision-making difficulties (Solove 2013, p. 1888).

Moreover, "little bits of innocuous data can say a lot in combination [...] it is virtually impossible for a person to make meaningful judgments about the costs and benefits of revealing certain data" (Solove 2013, p. 1890). This leads to situations in which "people consent to the collection, use, and disclosure of their personal data when it is not in their self-interest to do so" (Solove 2013, p. 1895).

Not only this: the long-term storage of big data means that situations may arise in which there is a desire to use it for purposes that may not be remotely connected to those for which it was gathered (and for which consent was given) in the first place. What alternative measures might be taken to ensure that *genuinely* informed consent is obtained from those who give up their data *for all the purposes to which it might be subsequently put*? Or are Barocas and Nissenbaum correct in their assertion that "it is impossible, even absurd to believe that notice and consent can fully specify the terms of interaction between data collector and data subject" (Barocas and Nissenbaum 2014, p. 66)?

Further issues surround the practice of anonymizing data in order to preserve the privacy of those who have supplied it. Preserving that anonymity turns out to be deeply problematic. In practice, when data has been anonymized (or "de-identified") it may well be possible to break that anonymity by triangulating between multiple datasets. For example, in the American context "knowing an individual's ZIP code localises that person to one in 30,000 (the average population of a ZIP code). Linking a ZIP code with a birthdate reduces the pool to approximately one in 80, while further connecting gender and year-of-birth are sufficient, on average, to uniquely specify an individual" (Koonin and Holland 2014, p. 146). In other words, if different anonymized databases which contain my ZIP code, birthday, gender and year of birth are linked, it is highly likely that I can be identified (for further examples of the ways in which data can be de-anonymized see (Porter 2008)). The ease with which de-anonymisation may be carried out leads Raley to observe that "anonymisation cannot and should not be considered a means of privacy protection" (Raley 2013, p. 128). It is hard to resist the conclusion of Barocas and Nissenbaum that "[p]rivacy and big data are simply incompatible" (Barocas and Nissenbaum 2014, p. 63).

3.2. Security

How is data to be kept intact, and safe from accidental or malicious threats? Strategies for dealing with such threats typically take the form "prevent, detect, respond, recover" (Landwehr 2014, p. 214). It is incumbent upon data handlers to use whatever technologies are feasible to prevent the exposure of data to degradation or cyber-attack. However, given that not every disaster may be foreseen, and that those with nefarious purposes will always be seeking new ways in which to circumvent security measures, it is equally important to have robust systems in place to detect such attacks when they are made, respond appropriately to them, and ensure that the system which has been attacked is restored to its pre-attack state—with, if necessary, appropriate new safeguards in place.

Whatever measures are taken, the frequent occurrence of news headlines concerning the hacking of computers and publication of confidential material suggests that the possibility of security breaches will always be a problem. If this is so, then we might wish to ask: to what extent are those whose personal data might be compromised by such breaches made aware of the risks to which they are exposing themselves?

3.3. Ownership

Are data property? If so, who owns data? Should it be the person to whom it relates, or the organisation which has gathered it? Should something akin to copyright protection apply to data, to prohibit its use or reproduction by parties who infringe ownership, and maximise the opportunity for those who own data to extract profit from it? It has been observed that "the prime driver of big data is not technological; it is financial and the promises of greater efficiencies and profits" (Kitchin 2014, p. 119). Is this inevitable? Or should data be part of an "information commons", which might be understood as "a body of knowledge and information that is available to anyone to use without the need to ask for or receive permission from another, providing any conditions placed on its use are respected" (cf. (Tavani 2013a, p. 256))?

To illustrate this, consider the data which is routinely stored whenever individuals give samples as part of a medical procedure. That data may be enormously important for the conduct of research into the aetiology and treatment of diseases, and it might be urged that it should be freely available to medical researchers for that purpose. However, it might also be possible to mine that information for data which is of great commercial significance, in the development of new drugs, for example, or in the provision of health insurance. Should the data be freely available, to assist the researchers? Or should it be treated as a commercially-sensitive asset, with restrictions placed on accessing it?

3.4. Regulating Commercial Use of Personal Data

This brings us to another important set of issues. If big data might be used in such a way as to turn a profit, then what regulation should exist around its use, in terms both of its commercial exploitation, and of trade in the data itself? Data may be commercially useful in a number of ways, in addition to those already mentioned. Information about a person's past purchases, recorded by a website which they use, or through their use of a store card, can be used to target advertising and encourage further expenditure by that person. Many people will doubtless find it helpful to be alerted to products which they might like, regarding which they might otherwise have been unaware.

However, given the potential for deriving financial gain from people's data, a whole industry has sprung up around the generation of data products which can be sold for profit. As Kitchin explains it:

> Data brokers (sometimes called data aggregators, consolidators or resellers) capture, gather together and repackage data into privately held data infrastructures for rent (for one-time use or use under licensing conditions) or re-sale on a for-profit basis [. . .] Data consolidation and re-sale, and associated data analysis and value-added services, are a multi-billion dollar industry, with vast quantities of data and derived information being rented, bought and sold daily across a variety of markets—retail, financial, health, tourism, logistics, business intelligence, real estate, private security, political polling, and so on (Kitchin 2014, p. 42).

Kitchin further notes that "At present, data brokers are generally largely unregulated and are not required by law to provide individuals access to the data held by them, nor are they obliged to correct errors relating to those individuals" (Kitchin 2014, p. 44). A report by the Canadian Internet Policy and Public Interest Clinic (CIPPIC) on the Canadian data brokerage industry makes the sobering judgment that "the increasing accumulation of personal data and consolidation of databases leaves individuals vulnerable to abuses by those with access to the data. Once released into the marketplace, personal data cannot be retrieved. Potential uses of this data are limited only by law and ethics" (CIPPIC 2006, p. 47).

The financially-motivated phenomenon of data brokerage has developed swiftly in response to market demands. It has yet to be properly held to account by either ethicists or legislators, and it would appear that the longer this situation continues, the more this business is likely to mushroom, and the harder it is likely to become to impose ethical or legal restrictions upon it. This is an area where the need for concerted action is becoming more acute by the day.

3.5. Surveillance

The possibility exists of forming detailed pictures of an individual around that person's digital footprint: not only their purchases of goods and services, but also their communications via phone and email, and their physical movements (if they carry a device which has Global Positioning System (GPS) enabled). This has been termed "dataveillance" (Raley 2013). There may be justification for such activity, in terms of monitoring the activities of individuals who may be thought to be terrorists or security risks: this thinking lies behind the UK's Investigatory Powers Act 2016 (widely dubbed the "snooper's charter"), which "requires web and phone companies to store everyone's web browsing histories for 12 months and give the police, security services, and official agencies unprecedented access to the data" (Travis 2016). However, such dataveillance may also be seen as an infringement of privacy and, as such, it raises the same suite of ethical problems identified under Sections 3.1–3.3 above.

3.6. Entrenching Unfairness

In the United States in particular, big data has found widespread use in such fields as policing, assessing people's suitability for loans or for jobs, and targeted advertising aimed at everything from selling products to enrolling people in university courses. As O'Neil has pointed out in a thought-provoking analysis (O'Neil 2016), this frequently has the net effect of reinforcing existing social inequalities. For example, if a person's poor credit record and home address in a poor neighbourhood are judged to make them unsuitable candidates for a job which they are seeking, or are used as justification for charging them higher insurance premiums, this has the net effect of denying them opportunities and reinforcing their existing poverty. It has long been recognised that access (or lack of it) to the benefits brought by computers has the potential to deepen social inequality: the application of big data to categorising and sorting people has the capacity to take unfairness to entirely new levels. As O'Neil puts it, "The poor are expected to remain poor forever and are treated accordingly—denied opportunities, jailed more often, and gouged for service and loans. It's inexorable, often hidden and beyond appeal, and unfair" (O'Neil 2016, p. 155).

3.7. Generation and Analysis of Data

A raft of issues exists around the gathering of data, its treatment and analysis, and the presentation of the results of such analysis by experts to those who have little understanding of the way those results have been obtained, but who may be making potentially far-reaching decisions based on them (recall that one task of the data scientist is to be a storyteller: that is, to formulate persuasive narratives accounting for the results of data analyses).

It might be thought that a big dataset just "is". However, consider the following ways in which it embodies particular values and biases. (a) Data do not just happen: they are generated by one process, and captured or selected for retention by another. A variety of factors may introduce bias into these processes. (b) "Raw" data is routinely "cleaned" prior to detailed analysis, and it has been noted that "decisions about how to handle missing data, impute missing values, remove outliers, transform variables, and perform other common data cleaning and analysis steps may be minimally documented. These decisions have a profound impact on findings, interpretation, reuse, and replication" (Borgman 2015, p. 27). (c) The particular tools used to analyse "cleaned" datasets may, themselves, have preconceptions embedded within them: "the algorithms used to process the data are imbued with particular values and contextualised within a particular scientific approach" (Kitchin 2014, p. 136). (d) Biases in the interpretation of data may lie not only in the tools analysts use, but in the analysts themselves. boyd and Crawford note that "researchers must be able to account for the biases in their interpretation of the data. To do so requires recognizing that one's identity and perspective informs one's analysis" (boyd and Crawford 2012, p. 668). (e) Additionally, Rosenberg observes that such individual bias is, itself, shaped by the context in which the individual is located: "from the beginning, data was a rhetorical concept [. . .] As a consequence, the meaning of data must always shift with argumentative strategy and context—and with the history of both" (Rosenberg 2013, p. 36).

It is crucial, then, to understand the biases built into the ways in which data are generated, the tools used to "clean" and analyse them, the influences acting on those undertaking the analysis, and the context in which the analysis in undertaken. As O'Neil and Schutt warn, "it is wrong to believe either that data is objective or that 'data speaks', and beware of people who say otherwise" (O'Neil and Schutt 2014, p. 25). However, to what extent is this kind of nuanced understanding pursued in practice? Commenting on a study of those engaged in data analytics, Kitchin notes: "Worryingly, those who 'let the data speak for themselves' [...] outnumber those best able to make sense of big data" (Kitchin 2014, p. 161). It would appear that the subtle and nuanced approach required of data scientists in practising their craft is still at an early stage of development.

Attention needs to be paid not only to the analysis of data, but to the presentation of the fruits of that analysis. This will frequently involve the production of visualisations of the data and of what it is telling us, through graphs, charts, diagrams, and so on; and these visualisations may, themselves, incorporate conscious or unconscious bias in those who have prepared them, which might encourage those to whom the data is being presented to read it in particular ways. As Gitelman and Jackson put it, "Data visualisation amplifies the rhetorical function of data, since different visualizations are differently effective, well or poorly designed, and all data sets can be multiply visualised and thereby differentially persuasive" (Gitelman and Jackson 2013, p. 12). Given that decisions affecting the lives of thousands of people may rest on such visualisations, it is crucial that those who devise them clearly understand their impact.

3.8. The End of Science?

A sensationalist response to the arrival of big data analysis has been to see it as effectively replacing science, as it has been practiced hitherto, with an entirely new approach to generating information about the world. In an article titled *"The End of Theory: Will the Data Deluge Makes the Scientific Method Obsolete?"*, Anderson has written that:

> This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behaviour, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves (Anderson 2008).

This kind of rhetoric strikes the present writer as both alarming and dangerous since, as we have seen, numbers emphatically do not "speak for themselves", but rather require careful explication and acknowledgment of the complexities and biases present in their collection, analysis, and presentation. It is to be hoped that more cautious voices will rapidly come to prevail when thinking about what big data can, in practice, achieve.

Even if that is so, however, data science may still present challenges to science as it has traditionally been understood. As an example of this, consider the much-discussed Google Flu Trends (GFT) project, which attempted to use data input to the Google search engine in order to predict outbreaks of flu in the United States (cf. (Fuller 2015)). Traditionally, scientific results are published in such a way that they are (at least in principle) reproducible; however, it was pointed out that publications relating to the GFT project withheld data and did not disclose the particular search terms on which the project was based. Presumably, commercial reasons lay behind this withholding of information. It, nevertheless, represents an approach to the publication of scientific research which is incompatible with traditional understandings of scientific practice.

4. Addressing the Issues

The issues identified here are complex and overlapping. How might they best be addressed? I suggest that there are, broadly, four ways in which this might happen, and I would urge that at least some of these ways of addressing the complex issues raised by big data can be informed by thinking, and by practical activity, which may come from theological and religious communities.

4.1. Technical Responses

Some of the concerns raised above may be addressed through technological development and innovation—that is, through the upgrading of hardware and software systems used to store and process big data. For example, as suggested earlier, some of the issues relating to security can best be dealt with by constant monitoring and development of infrastructure and of protocols designed to prevent accidental or malicious contamination of data, or the release of personal information. Since data science is a rapidly-changing and advancing field, it is likely that techniques for improving security will similarly change and develop—and it is equally likely that those who wish to counter that security will also change and develop their methods for doing so. Insofar as these technical responses are responses to unforeseen events, it is difficult, if not impossible, to plan ahead regarding them. Ultimately, though, those whose lives may be affected by technical failures and security breaches should, at the very least, be informed of the risks they are undertaking, and consent to those risks.

4.2. Legal Responses

A number of matters might be dealt with by legal and regulatory means. However, there are significant difficulties here, of which we may note three which are particularly acute. First, the rapidity with which big data, its analysis, and its commodification are developing means that action is required urgently if legislation is not always going to be several steps behind current practices, with the risk of its becoming obsolete at the moment it hits the statute-books. Second, there may be circumstances in which legislation already exists, but in practice it is either unfit for purpose or routinely ignored (or both). In such cases, the adaptation or enforcement of laws needs to be made a priority (cf. (O'Neil 2016, pp. 212 ff.)). Third, there is the problem of making legislation universally applicable. Data is not restricted by borders, but laws are enacted by nation-states which may vary greatly in the ideals which they wish to enshrine in those laws. Much of the discussion around the regulation of big data thus far has focussed on the U.S. context, but it should be noted that other, well-developed legal frameworks are operative elsewhere, such as those set out in the UK's Data Protection Act 1998 and the European Union's General Data Protection Regulation of 2016 (see, e.g., (Elias 2014)). Ideally, international regulation of big data gathering, usage, and trade is surely desirable. It may be possible for some international standards to be agreed, but it is likely that there will be a continuing need for coordination and cooperation between countries in producing arrangements for the cross-border policing of such issues. These may be significant (and significantly expensive) undertakings, but they are surely crucial.

4.3. Ethical Responses

The tools used by data scientists have not tended to prioritize ethical considerations. O'Neil has urged that "we have to explicitly embed better values into our algorithms, creating Big Data models that follow our ethical lead. Sometimes that will mean putting fairness ahead of profit" (O'Neil 2016, p. 204). Since much work with big data is profit-driven, ethical frameworks are required that can override financial considerations. How might these be generated?

Tavani has noted that many professional societies—including some related to the computing profession—have adopted codes of conduct governing the behaviour expected of their members (Tavani 2013a, p. 106 ff.). In the data science context, we may note that the Association of Internet Researchers has produced recommendations on ethical decision-making and internet research for its members (Association of Internet Researchers 2012). Guidelines and recommendations are a good start, but more is surely required to ensure that all data science practitioners act in ethically responsible ways.

Now, although a science based on the manipulation of data might appear "objective" to an outsider, we have seen that this is, in fact, very far from being the case, and that data scientists are required to use considerable personal skill and judgment in their work. This, perhaps, makes data science more akin to the practice of medicine than to that of a "hard science", like physics. This, in turn, suggests a possible way of addressing those issues noted above which relate specifically to the practice of data analysis. Doctors have a touchstone for ethical behaviour in the Hippocratic Oath (cf. (British Medical Association 2012, p. 887)), and it has been suggested that data scientists, too, might undertake a similar oath, holding them to particular ethical standards in the practice of their craft (O'Neil 2016, pp. 205–206; Fuller 2015, p. 581). This might involve an undertaking to ensure that their work is conducted with fairness to all the parties on whom it impacts, and that it is used to ends which promote human flourishing, rather than otherwise. Were such an oath to be devised, it might be an occasion for collaboration between data scientists, ethicists, and those from religious traditions in shaping the form it might take. Paul Ohm (Ohm 2014, pp. 108–9) has averred that "as we expand the reach and power and influence of data science, we must take steps to prevent harm, to ensure that this remains always a humanistic endeavour, and to help people preserve their power and autonomy". That might not be a bad set of objectives from which to start.

The adoption of a binding oath by data scientists might be thought of as a way in which many ethical concerns might be addressed. However, for this to be effective (a) the oath would have to be obligatory for all data science practitioners, and (b) sanctions or penalties would need to be applied to any practitioners found to be in it to be in breach of it. There are no indications at present that this is a likely scenario. If self-policing is ruled out, how else might issues raised by data science be addressed? What other bodies are in a position to raise and discuss them?

Many of the concerns raised by big data are fuelled by the ignorance of the general public regarding the ways in which data are used, stored, and traded. Given the many vested interests involved in the ownership and trading of data, and the likelihood that these will lobby against measures which would see a diminution of their profits (both real and potential), the addressing of this ignorance is a matter of some urgency. Education—making people more aware of the consequences of their giving away information about themselves—is clearly important. So, too, is counter-lobbying, which is likely to be required if legal measures are to put in place around big data issues. What fora exist which might address these educational and lobbying tasks? Dedicated pressure groups, such as those concerned with the privacy of citizens, will have an important part to play in addressing these issues, but these tend to involve relatively few people. How might more citizens become engaged with the new problems raised by big data?

4.4. Religious, Theological, and Hermeneutical Responses

Religious communities, such as churches, are surely in a position to play a significant part in raising awareness and encouraging discussion of these matters. Christianity, in common with many other faith traditions, places a high value on notions of accountability and fairness. Where such values are seen to be being flouted—for example, through keeping people ignorant of the consequences of their giving up data, or through the use of big data to reinforce social inequalities, or through the gathering, storage, and trade of personal data without effective consent, by unaccountable commercial organisations—the churches might speak out against such practices, and lobby for their restriction. Such church involvement need not be solely concerned with restrictions. There are very large benefits which the application of big data may bring to society, and the churches might, therefore, also lobby to ensure that such benefits are distributed amongst all citizens (cf. (Fuller 2016)). To give an example of such engagement in a different context: the Church of Scotland, through its Society, Religion, and Technology Project, addressed issues of public concern raised by the cloning of Dolly the sheep through setting up a group of academics from the biological, agricultural, and social sciences, together with theologians and ethicists, to explore bioethical questions raised by this advance. A subsequent publication made the fruits of their discussions widely available to the lay public (Bruce and Bruce 1998).

Such practical actions as these are likely to come from the organisational leadership of religious organisations like churches, but such bodies have further roles to play in that they are communities which allow issues to be discussed and "owned" by a broader public than that which is likely to engage with issues when they are presented in purely ethical terms. They can also constitute fora which enable an engagement with these important issues through both "head" and "heart". Ethical matters can elicit visceral, as well as intellectual, responses, and churches offer areas of engagement where both may be engaged effectively. Churches and other religious groupings can, thus, offer a "safe space" in which people's enthusiasms, hopes, and fears regarding big data might be discussed, and openness with regard to big data might be encouraged. As awareness of the questions surrounding big data increases, the need for such a space will inevitably become more and more acute.

In parallel with public action of this kind, religious engagement with issues raised by big data can also take place through raising theological concerns at a more academic level. It is important not to lose sight of the fact that data constitute just one part of a much greater picture. Individuals focussed on particular data-analytical tasks may lose sight of the human stories which lie behind their data, and of their work as anything other than the application of mathematical processes, directed towards abstract ends. Effectively, human beings are reduced to numbers, patterns, and trends. Here, insights from theological anthropology might offer a helpful corrective. It might be urged that every person is unique, complex, and formed by their relationships with others and with God. Moreover, people are intrinsically valuable, rather than having a value conferred upon them through the measurable parameters that can contribute to a large dataset. Theological insights such as these have a part to play alongside purely ethical critiques of big data. Both, alike, can serve the purpose of constantly reminding data scientists and data subjects alike of the "big picture" within which all of our endeavours are set, and by insisting that the goal of those endeavours should be the promotion of human flourishing. Both can also give voice to the imperative that big data be used justly, so that the gains which might be made through its use are shared as widely as possible, and benefit as many people as possible.

There is a further, very particular, skill which the theological community can offer to the practice of data science. It has been observed that "when the amount of data is sufficiently large, you can find almost anything you seek lurking somewhere within" (Berman 2013, p. 145). As noted above, data scientists are engaged in a complex interpretative exercise which involves recognition of the history and context of the data which they are analysing, the biases contained both in it and in the analytical techniques which are being used to explore it, their own inbuilt biases, both conscious and unconscious, and the complexities of the (quite possibly, commercial) context in which their work is being carried out. Complex interpretative exercises of this kind are familiar to those theologians who deal with the interpretation of texts, and they have developed a suite of hermeneutical skills to assist them in engaging with such interpretation. A dialogue between data scientists and theologians concerning hermeneutical practices could be an important way in which skills developed in the service of a religious tradition might also valuably inform practices within this developing new science—an idea I have developed more fully elsewhere (Fuller 2015, pp. 577–80), and such a dialogue is likely to be of considerable benefit to theologians, too (Fuller forthcoming).

5. Conclusions

The arrival of big data has already brought with it numerous questions that have yet to be properly addressed, and others will doubtless emerge as data science develops further. These questions are methodological, epistemological, and ethical, and they concern (inter alia) the ways in which data are collected, stored, interpreted, re-presented, and traded. A further complication is the speed with which data science is advancing, which means that (for example) the application of legal and ethical restrictions to the practice of that science will always risk being several steps behind the point that it has currently reached. There are indications that we are currently sleepwalking towards a situation in which the commercial exploitation of big data routinely increases social division, and renders privacy a thing of the past.

Many of the issues highlighted in this paper are complex, and may appear intractable. Addressing them will certainly involve the engagement of many parties—data scientists, lawyers, ethicists, politicians, and representatives of the business community and of the general public. There will also need to be an engagement of different jurisdictions, as solutions are sought which will command general consent, and which will have legally-binding international validity.

In addition to a purely "secular" treatment of these ethical issues, this paper argues that the engagement of religious communities and theologians has an important part to play. Religious communities may offer critiques based on particular sets of values which treat human beings as more than simply data points, and they may offer fora for discussions which may aid in the dissemination of information about data science, as well as an opportunity to critique it. In addition, there is enormous potential to be harnessed in bringing data scientists into dialogue with theologians, since the hermeneutical skills developed by the latter may have much to offer to the former.

The arrival of big data has brought with it concerns which are only starting to be appreciated and discussed. All resources—including those of religious communities—which can enable such appreciation and discussion are to be welcomed.

Conflicts of Interest: The authors declare no conflict of interest

References

Anderson, Chris. 2008. The End of Theory: Will the data deluge make the scientific method obsolete? *Edge*. June 30. Available online: https://www.edge.org/3rd_culture/anderson08/anderson08_index.html (accessed on 11 January 2017).

Association of Internet Researchers. 2012. Ethical Decision-Making and Internet Research. Available online: https://aoir.org/reports/ethics2.pdf (accessed on 17 January 2017).

Barbour, Ian G. 1992. *Ethics in an Age of Technology*. London: SCM Press.

Barocas, Solon, and Helen Nissenbaum. 2014. Big Data's End Run Around Anonymity and Consent. In *Privacy, Big Data and the Public Good: Frameworks for Engagement*. Edited by Julia Lane, Victoria Stodden, Stefan Bender and Helen Nissenbaum. New York: Cambridge University Press, pp. 44–75.

Berman, Jules J. 2013. *Principles of Big Data*. Amsterdam: Elsevier.

Borgman, Christine L. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge: MIT Press.

boyd, danah, and Kate Crawford. 2012. Critical Questions for Big Data. *Information, Communication and Society* 15: 662–75. [CrossRef]

British Medical Association. 2012. *Medical Ethics Today: The BMA's Handbook of Ethics and Law*. Oxford: Wiley-Blackwell.

Donald Bruce, and Ann Bruce, eds. 1998. *Engineering Genesis: The Ethics of Genetic Engineering in Non-Human Species*. London: Earthscan Publications Ltd.

Canadian Internet Policy, Public Interest Clinic (CIPPIC). 2006. On the Data Trail: How detailed information about you gets into the hands of organisations with whom you have no relationship. A Report on the Canadian Data Brokerage Industry. Available online: https://cippic.ca/sites/default/files/May1-06/DatabrokerReport.pdf (accessed on 11 January 2017).

Elias, Peter. 2014. A European Perspective on Research and Big Data Analysis. In *Privacy, Big Data and the Public Good: Frameworks for Engagement*. Edited by Julia Lane, Victoria Stodden, Stefan Bender and Helen Nissenbaum. New York: Cambridge University Press, pp. 173–91.

Fuller, Michael. 2015. Big Data: New science, new challenges, new dialogical opportunities. *Zygon* 50: 569–82. [CrossRef]

Fuller, Michael. 2016. Some Practical and Ethical Challenges Posed by Big Data. In *Embracing the Ivory Tower and Stained Glass Window: A Festschrift in Honor of Archbishop Antje Jackelen*. Edited by Jennifer Baldwin. Heidelberg: Springer International Publishing, pp. 119–27.

Fuller, Michael. Forthcoming; Boundless riches: Big Data, the Bible and Human Distinctiveness. In *Issues in Science and Theology: Are we special?* Edited by Dirk Evers, Michael Fuller, Anne Runehov and Knut-Willy Saether. Heidelberg: Springer International Publishing.

Gitelman, Lisa, and Virginia Jackson. 2013. Introduction. In *"Raw Data" is an Oxymoron*. Edited by Lisa Gitelman. Cambridge: MIT Press, pp. 1–14.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage Publications.

Koonin, Steven E., and Michael J. Holland. 2014. The Value of Big Data for Urban Science. In *Privacy, Big Data and the Public Good: Frameworks for Engagement*. Edited by Julia Lane, Victoria Stodden, Stefan Bender and Helen Nissenbaum. New York: Cambridge University Press, pp. 137–52.

Landwehr, Carl. 2014. Engineering Controls for Dealing with Big Data. In *Privacy, Big Data and the Public Good: Frameworks for Engagement*. Edited by Julia Lane, Victoria Stodden, Stefan Bender and Helen Nissenbaum. New York: Cambridge University Press, pp. 211–33.

Laney, Doug. 2001. 3D Data Management: Controlling Data Volume, Velocity, and Variety. META Group Report. Available online: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (accessed on 20 July 2016).

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. London: Allen Lane.

O'Neil, Cathy, and Rachel Schutt. 2014. *Doing Data Science*. Sebastopol: O'Reilly.

Ohm, Paul. 2014. General Principles for Data Use and Analysis. In *Privacy, Big Data and the Public Good: Frameworks for Engagement*. Edited by Julia Lane, Victoria Stodden, Stefan Bender and Helen Nissenbaum. New York: Cambridge University Press, pp. 96–111.

Porter, C. Christine. 2008. De-Identified Data and Third Party Data Mining: The risk of re-identification of personal information. *Shidler Journal of Law, Commerce and Technology* 5: 1–8.

Raley, Rita. 2013. Dataveillance and Countervaillance. In *"Raw Data" is an Oxymoron*. Edited by Lisa Gitelman. Cambridge: MIT Press, pp. 121–45.

Rosenberg, Daniel. 2013. Data before the Fact. In *"Raw Data" is an Oxymoron*. Edited by Lisa Gitelman. Cambridge: MIT Press, pp. 15–40.

Solove, Daniel J. 2013. Privacy management and the consent dilemma. *Harvard Law Review* 126: 1880–903.

Tavani, Herman T. 2013a. *Ethics and Technology: Controversies, Questions and Strategies for Ethical Computing*. Hoboken: Wiley.

Tavani, Herman T. 2013b. ICT Ethics bibliography 2012–2014: A select list of recent books. *Ethics and Information Technology* 15: 243–47. [CrossRef]

Travis, Alan. 2016. 'Snooper's Charter´ bill becomes law, extending UK state surveillance. *The Guardian*. November 29. Available online: https://www.theguardian.com/world/2016/nov/29/snoopers-charter-bill-becomes-law-extending-uk-state-surveillance (accessed on 11 January 2017).

Wilbanks, John. 2014. Portable Approaches to Informed Consent and Open Data. In *Privacy, Big Data and the Public Good: Frameworks for Engagement*. Edited by Julia Lane, Victoria Stodden, Stefan Bender and Helen Nissenbaum. New York: Cambridge University Press, pp. 234–52.