Monash University

From the SelectedWorks of Marcus R Wigan

Winter December 8, 2015

BigData - Can Virtue Ethics Play a Role?.pdf

Marcus R Wigan

Available at: https://works.bepress.com/mwigan/29/

The Ethics of big data: Relevance of Virtue Ethics

Abstract

Big Data is a term for masses of information that is usually heterogeneous, usually from multiple sources, in multiple formats and at a scale of at least terabytes, and often substantially larger. It may be a data stream, or an assemblage of exiting large, not necessarily homogeneous datasets; both often contain large personal data content and thus can invoke ethical issues.

As a result of rapid disintermediation of wide areas of the economy and daily life, and the growing data and information intensity that has both enabled this and is creating many fresh forms of Big Data on a real time basis, it is important to ensure that the implications are understood by the communities affected. This had not occurred until recently in the areas of government surveillance (Mathews & Tucker, 2014), and when it did had a massive impact across the world. Expectations were changed (See Fig.1) and the emergent power asymmetries emphasized. Concerns over the ethical and power implications are now reverberating, with Australia moving to consolidate ever stronger asymmetric information powers over the community (http://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=s969), and the term 'Snowden Effect' has now achieved currency (https://freesnowden.is/frequently-asked-questions/).

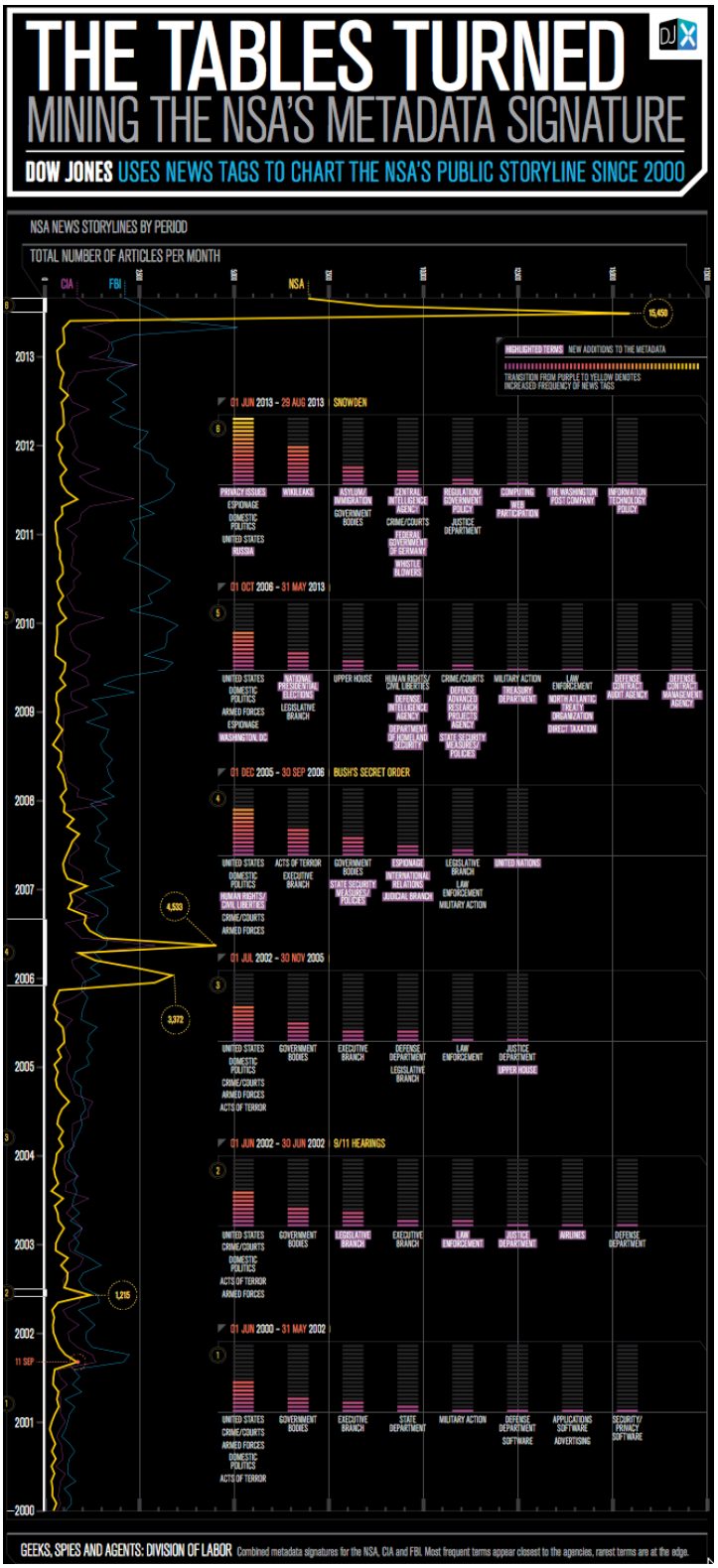

Fig.1.Order of magnitude changes in search behaviours (http://dowjones.com/pressroom/docs/DJ_NSAinfographic_v14.pdf,)

Most discussions about Big Data have been slanted towards the already evident economic or financial benefits of its successful application, with some work on privacy aspects. Generally there is still comparatively little usable discussion of the ethical issues involved in the struggles of science and marketing handling huge datasets, and little guidance for the professionals collecting or making use of it, or for policy and operational areas in and out of government embodying it into their capabilities and applications. Even agreements on the general principles that to be applied are far from being resolved: arguably a clear and present ethical quagmire and danger to the world community until this can be secured.

The Aristotelian (VI.13, 1144b1-17) concept of *phronesis* is essentially terms 'judgement'. It is indispensible for right conduct: i.e. turning good intentions into appropriate actions. Alternatively, 'doing well' has its own internal value, in the excellence of exercising the human intellect yet virtuous actions can and must also require phronesis. This provides a framework free of utilitarian tradeoffs between a variety of good and bad outcomes covering the design, the intent and the actions that Big Data and its supporting analytics enable. This provides a way to discuss the intent and virtue of appropriate actions, enabled by Big Data capacities.

Introduction

Big Data is a term for masses of information that is usually heterogeneous, usually from multiple sources, in multiple formats and at a scale of terabytes or above.

Most discussions about Big Data have been slanted towards the already

evident economics or financial benefits of its successful application, with some work on privacy aspects, but generally little discussion of the ethical issues involved for the professionals collecting or making use of it. A typical example is the recent oversight report of a selection of the current projects at Imperial College London (https://workspace.imperial.ac.uk/data-science/Public/2014.Big-Data-for-Better-Science_web-distribFINAL.pdf) makes no mention at all of ethics – nor are there any in the archives of the bulletin board associated by the Institute, prior to my own posting there. A more alarming lack of any mention of ethics is the major US report (Mellody, 2014) addressing teaching needs in the area by some of the most respected organisations in the USA. Two quotes emphasise the personal demands on participants, which should not allow them to avoid ethical issues.

> *"....data analysis process is highly interactive and iterative and requires the presence of a human being in the loop"* (p17), & *"…stressed the importance of provenance management—which is not often taught*" (p19)

The perspective that big data is simply 'more data', and only requires new techniques to allow better understanding or analysis clearly has considerable momentum.

The application of professional ethics is but one dimension, and an exploration of the virtue and utilitarian forms of ethical analysis is then made as they might, or should, be applicable.

The arguments are then summarised, and conclusions drawn, and recommendations made.

Mass Surveillance

Surveillance data is already overwhelmingly hug - and growing. Since Snowden the sheer scale and global reach has become part of public consciousness. The reactions are still continuing. Right wing political movements and Governments (Australia being one) publicly regard Snowden as a traitor and criminal, and are moving to consolidate the asymmetric information power that it gives the State, and the ever reducing civil rights that this incurs are regarded by this political movement as being of no great importance. This view is not shared by a substantial and increasing fraction of the community, who regard Snowden as a hero and an appropriate Nobel Peace Prize nominee.

It would be difficult to find a more universal example of an ethical divide. It embodies quite different perceptions of the duties of the State, the balance between autonomy and control, and the implementation strategies raise further questions that must be publicly debated and faced.

The most difficult issue is the collision between the Westphalian State view of state sovereignty (Beaulac, 2004; Caporaso, 2000) and the global citizen model that has emerged where many different trans-state actors now play an equal or greater role within almost every state. The concept that the (perhaps sole) core function of the State is security has become contested from a range of different viewpoints, not only citizens.

This is an example of disintermediation between the citizen and the actions of state security that emerged through incremental changes in globalization – followed by the huge series of leaps in information capture and scope that are now truly global and near comprehensive in coverage.

The language of disintermediation – all too familiar to businesses suffering radical change from the removal of middlemen in commercial processes – has not been applied to the surveillance domain, and its concepts of nation state security.

We argue that it is entirely appropriate, as the globalization of personal information, location connections and transactions changes the nation state as the basis of social values and ethical balances (through its national security powers) has already moved individuals into a global risk space where nation state concepts work very badly, and where nation states simply cannot exercise their expected national security functions without constraining or removing the very freedoms for individuals that they are charged with retaining at a survival level of ethical actions.

Unsurprisingly the command and control mechanisms, based on geolocation are now seeking to subject their citizens to greater restrictions (if their Constitutions permit[1]) as the nation state tries to retain its control of 'security'.

This is an emergent ethical conflict area that big data has not yet been generally associated with in public discussion, Yet, as it is a direct outcome of comprehensive mass data at the exascale, it is a critical field of Big Data impact on ethical issues of intrinsic worth of humans and autonomy (virtue ethics) and on utilitarian and consequentialist models of State actions for nations, which are still dominated by concepts of physical security.

[1] Australia has a dated and very limited Constitution bereft of any meaningful human rights in the relevant areas unlike US or European Union States, which have several types of applicable protections.

Correlations not Causality

The basic approach of Big Data is to use **all** of the available data, however inconsistent, incompatible in its metadata, and generally unverified and unverifiable in its content. The power of the widely used Open Source Apache Hadoop and Spark cluster computing software systems[2] is their capacity to handle these often-incoherent data assemblies. There are other major approaches, of which perhaps the most widely used are Bayesian methods and deep learning neural networks to extract, once again solely probabilistic, insights. Due to their Open Source nature there are no licensing barriers to access to such systems, however expensive the data access provisions and hardware might be!

This approach leads inevitably to using a wide range of inconsistent materials, and will contain errors, mis-specifications and gaps: will be in incompatible formats and structures (text, measurements, transactions etc.) as these are the characteristics of comprehensive data capture and are essentially unavoidable. Even the meanings and interpretations of many of the terms used in context or specifications will vary, as formal metadata specifications are unlikely to exist, let alone be coordinated across data type domains.

There are three reasons for this.

1. The diversity of data resources now accessible and usable in a single Big Data exploration have quite different conceptual structures (text notes, images, data readings etc)
2. The sheer scale of data pre-empts any realistic capacity to fill gaps,

[2] http://fortune.com/2015/09/25/apache-spark-survey/ ;https://bigdatauniversity.com/courses/spark-fundamentals/ and http://www.planetcassandra.org/getting-started-with-apache-spark-and-cassandra/ provide a concise clear exposition of what these are, and are given here as a basic guide to what they do

check for misreadings etc.

3. Increasing numbers of data streams are in real time, thereby making even current automated data cleaning processes unable to cope.

Making use of this growing torrent of multi petabyte-scale data sources (for a sense of perspective consider the entire administrative dataflow (shortly to be to be added to by the mass of actual medical records) of the medical system of an entire country, every motor vehicle movement captured by automated number plate recognition, every (poor) image recorded on driving licences – the flows of (poor quality) data extracted from CCTV records…. and that is simply a small subset of the substantially error prone dataflows already being captured)

A current ethically problematic example may help to appreciate the scale of these datasets and data transfer levels. Perhaps the biggest of all is the metadata retention requirements recently imposed on Australian ISPs and communications operators, these will require sufficient storage that the (clearly modest) $113million payment assessed for the ISPs to be able to hold this data for just two years that sheer size of this data capture. The illusive term 'metadata' was used deliberately by the Government to disguise the massive scale of data holding and the intrusiveness of it. Up until now the ISPs and telecommunication authorities held only enough data to be able to bill and audit their charges—and so retained only a tiny fraction of the micro level detail now mandated. However the confusion surrounding the actual levels of detail to be held was clearly intentionally introduced into the legislation by the Attorney General's drafters, to enable a broad span of demands for access and a micro level over time.

The limitation to two years still requires very large-scale storage, and it is likely that the security agencies will demand longer-term holdings for access to any specific two years, as they and the Government agencies have form in this regard. No provision was made for the very substantial security arrangements that will be needed for this mass community surveillance data base, no data breach requirements, and no penalty regime, in accord with the government's reluctance to date to legislate in this area.

The ethical struggles of the admirable Edward Snowden have made it very clear the levels of ethical strains on both implementers and administrators of this essential politico-legal Big Data initiative. The Nuremberg defence cannot be applied to such situations, and irrespective of security based penal government secrecy legislation, the personal accountability remains undeniable. Comparatively little attention has been paid to these ethical issues for the individuals involved, as an assumption that contract law and largely unaccountable legislative and administrative action will suffice. Ethical concerns were simply not considered.

We owe a huge debt to Edward Snowden for his courage and integrity. We will need whistleblowers more than ever in the era of exponentially expanding collection and cross-linking of Big Data. The tokenistic measures to 'protect' whistleblowers in the new era of total data capture are still simply window dressing, lack contestability – and indeed any verifiable credibility.

This makes the core ethical issues of Big Data analytics ever more difficult to contest.

The outcome of these gargantuan (and error prone) data flows is straightforward. To make use of them there are really only two domains

· Ex Ante (monitor for persons of interest: *aka* Surveillance)

· Ex Post (correlation & AI analyses to identify relationships of value)

The vast scale of the data can be reduced to correlations and projections based on these associations, but with little or no hope of a causal hypothesis being tested and applied. Any causal chains attributed are likely to be ex-post interpretations of the correlation outcomes.

The heterogeneous data segmentations for analysis and machine learning and other techniques used can handle the diversity of data quality and types, and there is a major distinction between sheer scale, and the rate of delivery of gargantuan data streams.

There and many sources of mass data, from historical sources as well as current standard collections of different types, but the new issues raised by large scale scientific data collections raise special issues.

The Large Hadron Collider data flows required one of the 12 major internet distribution nodes to be sited at CERN, and the frequent loss of the odd 200mb of data was completely inconsequential compared to successfully farming out to 50+ data centres around the work in real time as the flow could not be contained. The scale has steadily increased, so that it is now virtually impossible to provide experimental researchers with their own copies to take away of the complete set of data that they capture in a single experiment. The data streams will dwarf even this scale for the Square Kilometre Telescope Array; spread over two continents, when it comes on line in the early 2020s.

It is the sheer experimental research momentum of such projects, and others

in many other scientific fields, that have perhaps led to a comparative neglect of the ethical implications of the methods being developed, as the applications in these areas of science are not usually focused in an ethically problematic framework.

Big (Personal) Data

However, once personal data become part of the mix, as it does for the majority of fresh Big Data initiatives by Government, it all changes.

The ethical issues of national security are essentially a community ethical standard area – but once personal data is collected on a large scale the differences between surveillance/national security and ethical treatment of human beings become blurred, and each side appears to be almost invisible to the other. Depending on from which angle the questions and consequent vulnerabilities are approached. The crossover is the point after national security bodies suck up personal data and communications and make use of them. Most critical is the step that makes them available to other bodies in the state and commercial system.

All state surveillance bodies undertake both security and economic surveillance, and the latter is frequently shared with commercial parties for country specific competitive reasons (Kehly, Bankston, Greene, & Morgus, 2014). Ironically the Snowden revelations made it very clear the US companies in the Cloud Computing domain were likely to lose billions now that this has this became public knowledge. The trust deficiency levels have yet to be seriously addressed, although major US companies (Apple and Google for example) are now under pressure from the US NSA to stop introducing

encryption and other protection processes to their products, which they now see as a commercial advantage in their world-wide clientele.

The transparency is necessarily weak if provision to non-security agency bodies for economic purposes is made, raising ethical questions about the provision to some and not to all, let alone the ethics of covert commercial intelligence gathering once moved outside the national security frame. Ethical and equity questions are then raised in terms of selective provision to different parties – and accountability to whom? Certainly not the Attorney General's Department: perhaps not to Departments of Trade... These governance and accountability questions need to be more widely discussed.

The convergences of interest in this area in information and power asymmetries between political parties and major commercial enterprises are serious ethical questions in their own right.

There is little specific to Big Data to say about this, but inevitably the scale of detailed surveillance data is now in that domain, and so we cannot avoid the question and the issues.

The questions of accountability and governance will always arise when surveillance is involved, but the focus of the present paper is on the ethics of Big Data specifically. This is not in general about the provision of data or results – it is about the implications and nature of actions based on such provision, and the special features of Big Data in these processes.

Given the necessarily messy state data in the Big Data assemblies, the major outcome is the identification of significant associations between factors (correlations). This can lead to at least two different forms of actions

1. Efforts to determine a causal relationship or relationships
2. Actions based on probabilistic results using Predictive Analytics (the terms used to describe techniques to determine what to do NOW from a specific big data correlation).

Very different ethical implications arise from the two courses of action.

The first fits into a commonly accepted approach of establishing a causal chain of events, and could produce a chain of evidence if required if such a research process is successful. The ethical questions can then be handled – depending on the context – by university Ethics Committees or similar processes. If the results, after such an establishment of a causal relation are complete, then justifications for actions based on these correlations can then be slotted into a chain of evidence and rational explanation if and when required. This process can then take over from the uncertainties, inconsistencies, and downright errors in the raw Big Data resources used.

It is the second that raises most ethical problems, as all that emerges is a correlation, and thus simply a probability of a given action being appropriate but without any evidential chain to verify this is a form convincing in a Court as evidence. All an expert witness could do would be to explain the processes required to achieve the probable relationship or action prediction, and would have to acknowledge that the data was almost certainly full or errors, biases, missing points and inconsistencies.

This arises from using population-based analyses, applied in a generalised Bayesian sense to maximise the probability that a specific person is a person of interest, and thus to be targeted or otherwise restricted. The person concerned may then try to get some contestability by going to court, and thus

run into this fundamental divergence between civil law and intelligence logic.

While this might appear to be reasonable - it is also quite likely in cases where (for example) real time sensor data is involved that the correlations and predictive analytics actions would differ continuously – and all without a causal chain being possible to explain or justify the actions. National security actions must necessarily be based: the objective being to avoid major negative events, before they occur. So inherently national security intelligence must work with probabilities, and combine diverse sources to achieve the best and best-targeted result where preventative action can be directed.

This might be to debar a certain category of people from access to specific locations, to target individuals demonstrating a series of associated behaviours, or simply responding to the probability that an event will occur with a heightened probability – simply due to a statistical association with no verifiability.

Similar problems arise in simulation model validation, where the only way of validating them is to undertake a recent concept of a 'process validation' (Donelley, Thompson, & Wigan, 2012).

When actions based on Big Data correlations have a negative impact on people, as one would expect to occur with a probabilistically based action, then how these can be traced back in a responsible manner to an individual- and particularly how any errors introduced from the original data or the probabilistic estimation processes can be corrected and the problems that then arise from continued propagation of these 'findings' associated with an individual

Personal data is now of high value, and is bought and sold – as well as stolen, and so any such propagation is increasingly likely to occur. Personal record holdings become a two-tiered problem.

1. The holding of the data and its cumulative assembly
2. The inability to correct it by any party due to the probabilistic inference chains used to create the categorisations involved

This leaves the unverifiable characteristics attached to the category that is assigned, and then how this is then further propagated.

Examples of major domains where this can (has, and will) occur, and that are slowly coming into public focus include

- Medical data (PEHCR)
- Location based data (LBS, GPS)
- Government specifically focused data (MyGov)
- Transport data (Tolls, MyKi, ANPR)
- Energy data (Smart meters)
- Internet of things (IoT-see also smart meters and grids)
- Marketing data (including dynamic person dependent pricing)
- Banking data (international transfers, loan assessments etc)
- Social media (can be used to get accurate personality profiles etc)
- Political data (CityBuilder and other databases still not accountable in Australia under the Privacy Act, yet contain massive amounts of opinion-

based and unverifiable 'data' on voters and callers), and a most attractive data feed to Big Data analytics for commercial and political ends alike.

These are all separate domains but are all underpinned by several factors

- Data volume and heterogeneity is so large that errors are fully acceptable and indeed have to be
- The marshalling of the data analytics is to determine probabilities that lead to actions
- The actions are therefore founded on 'faulty' data, and inference- not on an auditable trail of evidence
- The outcomes of big data conclusions are therefore from the Intelligence background and values and culture.
- They are certainly not consistent with civil law, chains of evidence and verifiability

This bi-valued characteristic raises a wide range of ethical issues. In each domain (some of which are listed above) there are different professional ethical perspectives and standards… and these are already raising unresolved (and possibly unresolvable) ethical problems.

A fresh style of problem is appearing (and highlighted at Black Hat 2015 (Anon, 2015)), that of the Internet of Things. This is a term used to describe the billions of devices with their own IP address and communications capacity. These can be light bulbs to microsensors, to fridges to door locks. They are expected to dominate the internet within a few years, and will carry with them not only huge (terabyte-scale) data flows- but also great vulnerabilities both

for the people using them and for the community, as devices will then be hackable. The StuxNet virus (Kushner, 2013), which was aimed at embedded systems such as these, typical of Internet of Things devices, illustrates that then becomes possible - and indeed with the weaponisation of the Internet (Schneider, Lyle, & Murphy, 2015), it is highly probable that such risks will then become widespread.

The Black Hat group of hackers has made it widely known that there is a massive under-rating of these types of risks –and indeed of most security risks- by management in the organizations that they serve. This raises very real ethical and responsibility issues for the management as well as the technical implementers of such tools. The Black Hat group has taken an ethical stance in raising these questions but would still be subject to management direction were they required to create such malware. Would the Nuremberg defence be applicable? Certainly the ethical strains led Edward Snowden to act as a whistle blower simply by becoming aware of some of the applications being made of the surveillance software on which he was working.

An important Virtue Ethics stance available to those who understand the sheer scale of intrusiveness that IoT will inevitably bring is to participate in Standards before the issues become insurmountable. The IEEE Standard under development for IoT (IEEE Standards Association, InProgress) specifically recognises the different types of technical and privacy, safety, and security issues that are raised by the imminent arrival of literally many billions of communicating and unobtrusive tiny devices.

A useful premise, as a test question, is the examination of the various virtue

and utilitarian ethical frameworks. This is discussed later, and in principle the two approaches are the considerations of the intrinsic value of a person in a non-metric sense, and the valuation of attributes of a decision in a specifically metric sense. Utilitarianism considers all aspects are tradable ('have utility' in the microeconomic sense): virtue ethics does not; either for the subjects of the ethical consideration or the individual making the ethically influenced decisions themselves.

It is also a valuable research framework is to contrast intelligence and civil law frameworks and the collisions already evident between them in operational use today, and the ethical issues raised (and rising) as these collisions become more numerous.

It is useful to explore how Big Data is being perceived, legally handled, and how ethical questions are seen to be within the perceived domain of different professions

While there are clearly issues that are unlikely to be readily resolved, some of these will be amendable to ethical resolution and this research question is examined from different angles

Professional ethics demand that individuals making decisions with an ethical component meet the 'ethics' of the relevant profession.

This might appear to be a straightforward process, but it is far from that. One example is the divergence between different professional organisations' required actions when an ethical issue arises. Consider a registered psychologist finding in a pre recruitment scan that a party already well along

the selection process has a psychological problem disclosed as part of the process. This would not usually be disclosable to the employer of the psychologist. But if the person was a Human Relations Society member and subject to the ethics of that body, then such a disclosure would have to be made to the employer. In either case actions could be taken without disclosure of the reasons for the discontinuation of a selection, as this example is posed. But what if a person was a member of both societies?

These issues might appear to be far form Big Data, however, the extensive and increasing use of social media scanning, and analytics methods to build a probabilistic picture of likely attributes of the selectee from these sources led to a high probability that the interviewee was likely to not stay (even if a very good prospect on all other criteria), then what decision would the HR/Psychology practitioner make?

What duties to record the analytics assessment? What duties to disclose to the party affected?

As it would be a probabilistic assessment, there would appear to be no duty to advise the subject in either case. And as the analytics would have been completed as a business task the disclosure to the employer has already been made.

Information and Virtue Ethics in Big Data Issues

The virtue ethics stance suggests that intrinsic values matter, and has often been criticised for not allowing actions to be assessed in a manner that can be

communicated and understood is the same way that consequentialist arguments from utilitarian perspectives can be expressed (Russell, 2009).

Intellectual property has underpinned many of the tensions that now exist in Big Data domains and are contingent of its (assumed) rights of ownership and handling. Once the basic principle that there is a critical mass of data about a person that becomes in some sense that person, moving directly into the ethical sphere (Wigan, 2010), raising political resistance from the broad citizenry.

In the other direction, we have the handling of such personal information as an asset of rapidly increasing value of great interest (and value) to intellectual property rights holders (in many cases these are quite distinct from and do not include the original IPR creators).

It is still not widely appreciated that the intangible (IPR etc.) far outweigh the tangible in the valuation of goods and services worldwide, and so the importance of such agreements – preferably on a more ethical foundation than seen to date - is undeniable.

> *"We should acknowledge the increasingly vital importance of the framework of implicit expectations, attitudes, and practices that can facilitate and promote morally good decisions and actions." (Floridi, 2015)*

This basic mismatch continues in a heighted (and equally secret) mode for the twin 'evil children' of ACTA and SOPA. The Trans Atlantic and Trans Pacific Partnership regional trade processes (TPPA, TTIP). While US industry and the

US Trade Representative have complete access to the developing drafts and negotiators, while citizens –and indeed businesses – in all the other countries involved do not.

As in ACTA and SOPA, WikiLeaks and whistle-blowers was the sole source of material information. Dummy 'consultations' were of course undertaken to provide face validity of the processes, but nothing of substance was ever provided: only a forum for expressing concerns without feedback[3].

The moral basis for this behaviour is purely utilitarian, and expressed in terms of 'you cannot negotiate in *'real-politick'* without secrecy', while the near universal impacts on private values and lives are not regarded as being of sufficient value to enter this calculus.

A virtue ethic stance would challenge this, as it imputes a non-negotiable and non-tradeable intrinsic value to humanity. However, like so much in the information age of ever growing and more cross-linked data, the assets of such pinpoint monitoring (and of course eventually control, as long as it remains asymmetric) are disregarded.

This imbalance has caused major public protests in Europe, where the IPR regime and privacy trade offs are more widely established and sustained by Human Rights legislation (sadly lacking in Australia). This indicates not only that the political and trade processes are not working (the breakdown of the TTP negotiations in Maui in July 2015 illustrate this) but also that the political

[3] Based on my personal attendance at the formal TPP consultation sessions held by DFAT

processes are attracting under greater scrutiny and engendering ever-declining levels of public trust.

The quarantining of access to the negotiation process for TPP/TTIP and similar business-driven initiatives is simply due to the lack of balance in their formation: the moral and ethical requirement that people be party to the controls, surveillance, and limitations of freedoms that then follow demand that such processes be open. This information asymmetry is a direct result of the corporate/political axis denying this participation.

As Branscomb foresaw very early on, precisely when the US National Intelligent Transport Initiative and the National Communications initiatives suddenly merged during a meeting at Harvard when the confluence of high video bandwidth form Its was seen as the infrastructure to economically expand broadband access, there were several warning notes voiced, not only by Branscomb, but others (Wigan, 1996), on the basis that what is now termed Big Data and easy to exploit would raise serious moral and ethical questions..:

"*We will build the kind of legal information infrastructure that we want and need. If it does not reflect the wants and needs or our new electronic communities, it will be because we did not participate in the process to make it reflect our desires, and hopes, and expectations*" (Branscomb, 1994)

A relevant point made by Floridi is that the physical and digital domains of existence and property now 'require separate treatment'. This flies against the International Accounting Standards for intangibles (http://cpaclass.com/gaap/sfas/gaap-sfas-142.htm). Intangibles must now be

brought to account, and also represent around 80% of developed country stock market valuations some time ago. The stock market valuations of Google, eBay, and Facebook alone ($600bn) compared to their net asset ($125bn) demonstrates the huge valuations attached to their Big Data holdings (Monga, 2014), irrespective of the FASB guidelines!

This further emphasises the business/utilitarian single dimension ethical and moral stances that underpin such formalised global agreements, and are – in an unfocussed manner – raising virtue ethics values. While Floridi asserts a need to separate these physical and intangible aspects of property (specifically Intellectual Property Rights), he also - confusingly – appeals to his terms '*infosphere'* which is largely a relabelled partial recognition of the classical assertion of human existence as part of a '*noosphere'* (Teilhard de Chardin & Tr: Bernard Wall, 1959; Teilhard de Chardin & Tr: Norman Denny, 1964). Floridi also asserts that Information Ethics is an essential new all encompassing ethical framework (a form of meta-ethics), necessary to address the issues of the information age.

But while this is critiqued by Doyle (Doyle, 2010) to show that it is still consistent with utilitarianism in most cases, other arguments (Ess, 2009) take the debate somewhat further.

The framework used by Nietzsche (Martin, 2006), of a master-slave bifurcation of ethical and moral values is highly relevant to Big Data in particular, as the asymmetries of information that Big Data access and analytics provide are very much divided between the ethics applicable to those making use of it – and those who are subject to their results. Nietzsche's arguments that

different criteria apply may or may not be defensible on a variety moral grounds – but certainly ring true in terms of the power relationships involved, and are not inconsistent with the demands of virtue ethics (Russell, 2009). Hartman (Hartmann, 2013) takes a situationist perspective for business ethics, which has much in common with these various explorations of applied virtue ethics. In each of the cases and theories and critiques, a differentiation between persons virtue and the situation in which it is judged emerges: practical ethics: *phronesis* (often summarised- somewhat simplistically – as 'practical ethics') emerges in almost every case intact - with its implied situational (and in our argument potential power) differentiations.

Our position is that the very differentiation that underpins each of these stances is the power relationship that Big Data creates, and this aspect has to date not been addressed by any of the theorists addressed to date.

The appropriate framing of the Virtue Ethics application to issues of Big Data and its exploitation arise in the actions chosen. As these will depend on the context and situational status of the decision, indeterminacy appears to be a necessary aspect of all parties in a given action acting virtuously. In this we concur with Swanton (Swanton, 2003), who further observes that the reasons considered may not be good reasons when the judgement of 'good' is made by third parties.

We can conclude from this philosophical section that Big Data issues can be treated using Virtue Ethics approaches, but that these may not be a unique guide to action in any specific case, even if virtuous methods and considerations have formed the basis of the decision to act.

A good person may not make good decisions, even in good faith. This is hardly surprising, as Big Data issues are complex and multilayered, and whether a virtue ethics or a utilitarian calculus is used, the same will apply in the lack of complete information.

A number of such multilayered and conflicting goal issues will be considered in the next section.

Ethical Conflict Examples

One of the emerging risks to individuals is the rising fraction of company's assets that are intangibles. This growth was evident as long ago as the early 2000's, well before the recent developments in the value of Big Data and the abilities to exploit it. Computer information was then more company records and holdings –this has changed drastically in scale and scope in the last decade – and the personal impacts are now far greater. The concomitant National Security aspects of agencies and their subcontractors, who now scoop up virtually all communications and retain them indefinitely at the Exabyte scale, will be addressed in a later section.

Invisible assets have been indicated in FASB accounting standards discussions mainly in terms of Human Capital and Patents and similar intellectual property denumerable assets. The rise in personal data as a key wealth-creating asset has been illustrated by the salutary astronomic growth in stock market value of Google and Facebook, Linked In and Twitter, as their major assets are essentially personal data and personal communications.

In view of the lack of contestability of this ownership (reaffirmed by the less than transparent and ever changing terms of use) and the extremely large revenues secured from what is essentially simply the sale of personal data, it is now quite clear that accounting standards (which define only 'invisible assets' but still require them to be brought to account, certainly at the closure of affirm) must move to value the personal data held as major assets.

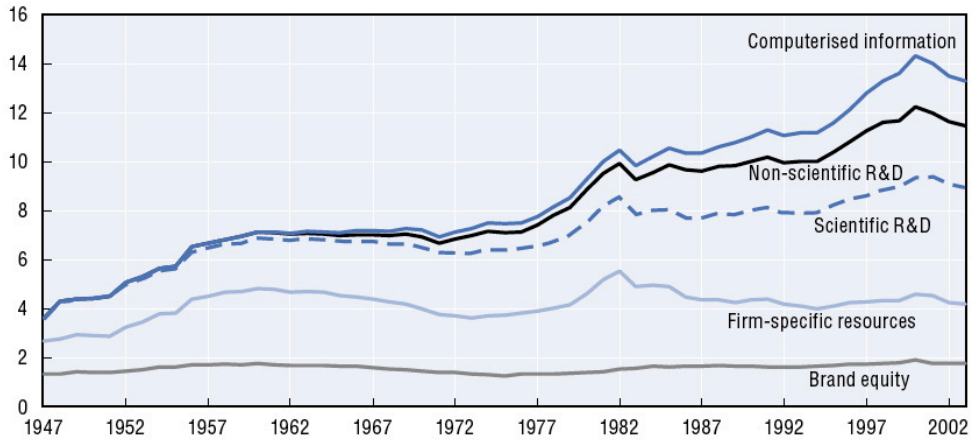

Fig. 2. Intangible investments in the USA 1947-2003 as % of business output (Source:(Corrado, 2005))

The more that they can be bought and sold the higher this value will be. Currently (at least as of the FASB statement of 2001 http://www.fasb.org/summary/stsum142.shtml) goodwill is the major issue, although the rising value of all forms of intangibles was explicitly recognised even then.

This does not raise any legal problems in terms of asset ownership, as we have no rights to the tokens that define our own legal identities. This might seem surprising, even pictures taken even in professional sitting for which you have paid, pictures from CCTV, credit cards, charge cards, etc. None of

these are your property, either individually or in aggregate. As a result and there is no way of way of contesting this situation under present Intellectual Property law on the basis of our having any ownership of our digital selves, other than by the law of defamation, and the first tentative green shoots of a privacy tort. However these apply only to suppression of availability of information held - and has no effect on its ownership. The ethical conflicts remain, and grow as the asset valuations in the sharemarket rely more and more on mass data, especially personal data.

However the politics of this imbalance in power have led to demands for the Right to Be Forgotten (European Union), as well as reputational and privacy torts are emerging in a number of countries, and comprise the first fine flush of some form of contestability of ourselves as 'business assets'.

Until these are resolved the ethical aspects are visible but not easily expressed or handled. However there is one situation which is already beginning to focus the attention of most parties, business, government and individuals: i.e. is the securing of personal data (especially of a scale that may be termed Big Data) when a company folds. And the liquidator must realize the assets as well as possible for the debtors for whom the liquidator is acting.

This places the ethical choices to be made by the liquidator in the firing line, both at a personal and at business ethical levels. Both Virtue Ethics (in terms of the personal decision as well as the valuation of the treatment of the people who are described by the Big Data set) and utilitarian ethics or maximises benefit while minimizing harms. And the business ethic of maximising realised value. These are difficult questions for both individual

decision makers and the subjects of the data. However the balance of the business ethos and the virtue ethical considerations of the liquidator are in clear conflict.

There is as yet no guidance on this from the Courts. But as de-anonymisation is now so straightforward, and so much selling identifiable individual data raises the value, the issue is now inescapable.

The move to intangible assets and information as a prime (if not the prime) source of 21C capital has a further complication beyond those previously considered (Contractor, 2001). As large scale data is now generally held and accessed using Cloud techniques, the location of such data is not longer clearly associated with any particular country or legal system, and it is common for extra territorial assertions of rights over such data held in these indeterminate locations to be made, not only by the USA (who has a long record of asserting extra-territorial applications of its own local laws) but now also by the European Community as the impacts on its citizens become more and more problematic.

The issues of global taxation domains and the treatments of transfer pricing are inextricably tied up in these issues, although this has not yet been widely recognised as an ethical issue: more one of revenue loss and eminent domain disputes (Boos, 2003). It is particularly important for Big Data applications involving personal data (or data that can at one remove be linked up to personal data (precisely or probabilistically – which introduces yet another complex dimension). There are various regional legal domains each of which apply stringent rules to verification of data held, such as the UK Data

Registration Act (https://ico.org.uk/about-the-ico/what-we-do/register-of-data-controllers/), but the application even of this Act to probabilistically associated data is effectively moot or inapplicable.

So where does this leave analysts and business operators working in this fuzzy globalised domain and looking for ethical guidance? Whose legal compliance regimes do they apply? That applicable to the location of the data source (which itself may already be stored spread over several different countries). The location of the application of the analysis? These too are often done over multiple global domains; finally, perhaps the location of the application of the results? This at least has a referral to decisions made by people who may be called to account.

The last is perhaps appears to be the most clear cut - but is the case of multinational companies almost as difficult to define as to handle. Legal systems and jurisdiction shopping is now well-established practice of multinational companies and the practical ethics of this (legal) practice are now user increasing pressure at a global level within OECD peak considerations. Simply holding personal Big Data in the UK must be registered under criminal penalties, so at what point in a Cloud held petabyte-scale distributed data holding does this become applicable? And if jurisdiction shopping is used, is this ethical? The debates over the (Weak) protections for data transfer between the US and the EU (Safe Harbour) are under increasing pressure[4] for imbalances of rights for the individuals whose data is affected. The indeterminate nature of 'location' in large-scale (Big Data) cloud

[4] As this paper was completed, the EU Ruled that the US-EU safe harbor was inapplicable: a determination long anticipated, and now formally in place.

databases is therefore a source on serious concern on many different types of fronts.

So where does this leave analysts, business operations managers, and marketers apply the results of a Big Data Analysis? In ethical terms there are no clear guidelines, certainly none that apply in all areas of a globalised data source, and a wide range of risks, both personal and commercial in the ethical domain. Is the plausible deniability of indeterminate location within a distributed cloud a sufficient legal protection? And if it is, is it ethical for an individual to take this stance? Or a business? Is it ethical for an individual to collude with this smearing of accountability? Once again, utilitarian minimisation of harms is far easier to handle that the ethical bases of individuals considering their own worth and those of the individuals affected. Virtue ethics raises the profile of such decisions in opposition to a simple utilitarian stance.

Virtue ethics applies to individuals' own choices to undertake specific analyses, but if they are kept ignorant of the application country or domain, does this help?

Chains of Professional Responsibilities

The complexity of the different threads that connect different sources of measured and derived data and their utilisation obscures the ethical responsibilities that should or could apply at any particular point in the chain.

This can be examined by exploring such chains, commenting on each of the different stages at which an ethical decision could or should be invoked, and testing if the decision makers have responsibilities that might or might not (depending on the organisational fragmentation of responsibilities and accountabilities) invoke a personal ethical decision to proceed. In general such segmented structures allow diffusion of responsibilities, and often when they become difficult, are further offset to consultants.

Consider a Big Data assembly from public transport ticketing. Here the data is collected at specific locations and times along each route and trip and mode change made by the holder.

The volume of such data is substantial in a city such as Melbourne, and in this particular case the system (MyKi) has proved to be subject to errors and failures still being worked out years after it was made live.

This unreliability in itself raises a series of ethical questions, as recent events have invoked pro bono barristers to intervene to assist those recorded in checks by enforcement officers that appeared to show that the activation had not been done for a specific journey.

However if using the full-scale database correlates patterns of timings, locations over time, then there is a clear basis for accepting that the database contains errors. Accessing these patterns can provide insights into the propensity of particular subgroups and their behaviours, which can be very helpful in indentifying demand patters. They also provide pinpointed data on the key locations of users at any given time, a very useful law enforcement

process both for tracking persons of interest and providing evidentiary information.

The identification of the likely locations and timing are probabilistic, and need be no more than that. Errors, even a substantial number, will not materially change these values for an aggregated subgroup. However providing evidence for a prosecution, in a system with known unreliability aspects, is more problematic.
This contract between mass data correlations and individual tracking, identification, and of course retrospective surveillance is a different set of applications, not the primary goal of a charging system. The core accounting requirement is to ensure that a proper charge is levied, but the

Chains of Ethical Responsibility

A major problem with Big Data is that the various aspects of ethical decision-making can so easily be offset to others in the chain.

An example makes this plain, let us choose a marketing example.

A retail shop installs a real time high resolution video tracking system, mainly for security, but he feeds are in real time and are acquired (let us postpone how) by a third party who combines them with many others.

The cumulative total is a large-scale multicamera data feed in real time… so what is done with it?

The associations with products on the shelves are captured from the high-resolution video, and also the faces of the people involved. The latter is simply a sod effect of the processing around the shop. This data is not all that reliable and is perhaps 30% inaccurate in the matches made even within the single shop. This data is used primarily for product positioning in the various shops and proves to be usable to determine the height and placement of a subset of products, enhancing the probability of their being picked up simply by proper placement

This is all very satisfactory to the retailer, who tunes his shop and product layouts… so far so good.

Now however there is a crime scene nearby and the police acquire the video records: two things happen, as the video records are not kept quite long enough, only the traces and face signatures are retained (and so cannot be verified), however this is enough for the police to claim that specific individuals are the people of interest. So all these candidate faces/ people/ biometric signatures go into the police working databases if these were assessed as probable offenders as distinct from loosely associated by unreliable facial recognition software. No appropriate metadata is recorded with this automated process.

The culprit is never found or charged, but these records remain as ‘persons of interest’ and are harvested by the Federal police.

Later on profiling software at an airport pulls one of them up and, as the crime scene was coded by the police as a possible terrorist related event, they

lose their passport and ability to travel outside the country. Having been tagged for action this person then becomes a high profile item in the various intelligence databases. These are 'leaked' to the local police, and cause a failure to pass a working with children/police record check. He she then loses his/her job

If this were not such a plausible chain of events it would be easier and more comfortable to assess: this mix of surveillance and civil law domains, this mix of probability–based allocation to one or more specific classifications and black letter regulations is already occurring.

The additional problem is that due to the intelligence aspects, no contestability is possible for the party involved, who in turn has no opportunity to even discover that he/she is supposed to have done or not done to trigger this disastrous chain of events.

No FoI request, no legal investigation, in fact there is no means of either discovering what has occurred or correcting it (assuming that it was a judgement based on an error of fact i.e. error by the image processing software, as distinct from an error of judgment, i.e. police tagging the event and image probabilistic identification to there being a terrorist association)

The next issue that occurred in this scenario was of a leak from the police database to a political party (who had targeted the individual as a political threat to their own preferred candidate) who included this assumed association to the other option based materials in their CityBuilder political

database. This database is in turn exempt from the Privacy Act (against the recommendations of the ALRC).

Who has ethical responsibilities in this chain? Anyone? Does this mean that we need a social ethical framework on top of personal ethics? Probably, but this treads on human rights conventions and has no place as yet in Australian Constitution, and so is a fragile thread, easily over ruled by administrative underfunding (as has occurred with the OAIC recently) or simple change in Parliamentary Act - or even, as is the case in many areas of national security, by regulation or Ministerial direction.

This might appear to be a long chain – were it not for the fact that every step on this chain has already occurred (albeit not all in the same event), and in situations where the scale of the data handling were orders of magnitude smaller that those now becoming routinely available from crowd scanning, police body cameras, and high resolution CCTV sources.

The interesting ethical issues arise more at the design stage than at the application stage, as the chain of events involves bodies and people with very different powers and purview.

One argument would be that probabilistic facia identification invokes an ethical concern for the integrity of the people subjected to it. Another would be that this decision is pre empted by the legal framework that applies to images in public laces. It is the combination of these capacities that move this from on occasional scarce event to a design criterion for surveillance where ethical issues in participation then clearly arise.

Once again, national security is a key frame for exempting separately questionable issues and actions from consideration in the larger interests of the community.

So where does the ethical concern arise?

Essentially at the point where intelligence, options, and records collated for putative National Security objectives become part of the civil law mechanisms, thereby bypassing the normal protections for civil society.

An additional difficulty in assigning ethical responsibilities comes from the legal frameworks criminalising any form of disclosure by the individual and his/her legal representation assuming that he/she actually discovered somehow that they were the subject of such an intelligence assessment action at all).

Here we have the basic conflicts between ethical decisions of people within an organisation and organisational goals and standards. Also the ethical concerns can so easily be transferred down the chain to others, so the ethical aspects are not salient.

Moving back one stage, we then have to consider the designers and implementers of the processes and software. The process design in essentially implementing legal operational requirements at each stage (until leaks or other dysfunctions are introduced - and these do raise ethical questions), but the software designers and implementers are not immune. This aspect raises

both practical and ethical risks: practical as the quality of the work is a professional responsibility and management risk; ethical as the levels of design precautions for constraints on the use or otherwise of the analysis systems themselves invoke difficult questions of design.

As so much Big Data analytical processes involve algorithms and deductive software, the ethical issues of software design to handle the basic data are somewhat overshadowed by the applications of the algorithms, which then create "data" themselves. The algorithmic design and testing then raises their own ethical questions.

Inherent Power Imbalance Issues

Big Data exemplifies the imbalances in information power very effectively. Not only can the holder of these massive data sets know exactly what is in them, but also their analyses are their own property. Yet the subjects have in general no knowledge, let alone control of participation in ether the holding, ownership of analyses made of them. As these large data collections where they include any items pertaining to individuals or groups, are essentially non-contestable by their subjects in ownership, use and exploitation, the conditions for an information power imbalance are clearly met.

Nietzsche offers and alternative approach to Virtue Ethics that aligns with business orientations (Martin, 2006), rather than humanist perspectives on intrinsic value of humanity. It pivots on an interpretation of 'good and bad in a manner that emphasises Master and Slave perspectives, arguing that Good is equated to success, and that only slaves have to balance lack of success with

an intrinsic separation of actions and personal merit. Clancey further argues that there is something unhealthy about this, as it derives from Nietzsche's assertion that 'basic human drives demand expression in some form', and if this has to be internalised then they will make us sick. The Master stance can of course fully discharge such pressures and the pressure to emulate the Master morality is increased. The merging of morality and actions that underpins Nietzsche's perspective and inherently his stance requires rankings (If not direct tradeoffs, as would be required in a utilitarian standpoint). Ironically this line of value-based argument still leads to clear and recognisable Virtues, with honesty and truth at the top - but truthfulness only as an internal honesty, and not one extended to ones enemies or opponents. Nietzsche offers many interesting perspectives:

> *'His writing resembles nothing in analytic philosophy journals, and it is no surprise that analytic philosophers in particular have had a difficult time appreciating his thought. But his views are clever and subtle, and grappling with them repays the reader manyfold' (Hales, 2000)*

One such is that

"*Life is will to power… It follows that power is the ultimate "value" sought by actors; all other values, such as causality, morality, logic, and even truth, are less important*" (Aspers, 2007)

This has a striking similarity to managing mass personal data without disclosing the outcomes to the subjects - but Nietzsche does demand generosity to the vanquished. There is still a steady stream of debate about

Nietzcshean ethics, and the inherent imbalances that they have embedded. Some argue that is it consistent, or at can coexist with, democracy as we see it from a liberal stance- but this is a somewhat strained argument. As Nietzsche offers a workable basis for business and government use of Big Data without much accountability, it is important to recognise that there are arguments to support this power differential and still remain within at least one school of Virtue Ethics (Daigle, 2006),

What would a 'good person' do?

The underlying premise of Virtue ethics ties back the ethical decisions to the decisions or actions of a good person. There are two ways that this can be considered:

1. **Avoidance**. By layering decisions in business, private life or policy such that the fine divisions between layers of responsibility are not treated as whole allows distancing from the test of a good person, this can be overt, as in the Challenger Shuttle case where an engineer was subject to a binding code of professional ethics, but the manager who made the decision was not.
2. **Covert**, where organisational professional ethics (usually procedural) are used as a way of deflecting the personal decision and thus the worth or otherwise of the person making the decision. Both techniques are widely observable as being in use, and are readily recognisable in professional society 'codes of ethics' as essentially risk avoidance strategies. Raised later here, in a slightly different context by.

Less obvious is that the use of one or both of these techniques may be organisationally driven as risk reduction strategies, allowing formal projections of 'ethical behaviour' while still being able to undertake unethical courses of action, as the individual responsibilities are so finely sliced and directed in different directions.

The area of Big Data presents special problems for personal ethical decisions, as the twin forces of organisations being forced to invest heavily in securing perceptions of 'trust' and 'ethical behaviour', while making extensive use of cumulatively highly personal data that most users are not even aware that they hold, secure or purchase. Efforts to 'anonymise' such assemblies of data are essentially doomed to failure unless very extensive measures are taken, and even then it is increasingly doubtful that de-anonymising of data in increasingly extensive combined data sets can really be resisted. It takes only a few items of data to re-identify the most carefully anonymised individual records. The sheer cumulative weight of the value of personal data to organisations has already been documented earlier, and shows the level of pressure on organisations to exploit sch data to the limits of the law. There are few ethical constraints to this process, other than the personal moralities and personal ethical decisions of individuals trapped in the entrails of major government, corporate, or political organisations. These are not always whistleblower issues, but more the cumulative daily operational decisions within the scope on individuals to consider and take.

The protections provided by slicing the layers of responsibility and roles in undertaking big data analytics, authorising their use, securing the assembled data, and assessing the individual impacts on individuals is in general placed

outside any single person or profession. Management has indeed no generally accepted and endorsed professional society ethics, and professions of organisational virtue are not in general credible (e.g. ComBank financial planning scandals). Analyses of professional ethics from a virtue ethics standpoint are almost all focussed on the legal and medical professions (Oakley & Cocking, 2001), and few on general managerial roles.

How do these slice and dice responsibility levels protect organisations and their members from having to make (or rather to take responsibility for) ethical decisions on Big Data and analytics?

The diversity and scale of the Big Data assemblies is such that errors and incompatibilities are unavoidable. Verified data in one data set will not survive the lack of cross matching and verification for its use with another, different, and differently defined data set. So fully verifying data is simply not possible.

Consequently the harms that might be done to an individual by inaccurate identification of certain characteristics that make them discriminated for treatment under analytics by allocating them to group of special interest (or price discrimination) – albeit probabilistically – cannot be audited, corrected or taken responsibility for as the usage lies in general outside the analytic domain.

The utilitarian perspective is to rate the weight of probable harms against the likely benefits, and to accept that such tradeoffs are intrinsically valid. The difficulties come when assessing the risks to whom in an environment where

risk offsets (in the organization) is a key corporate and managerial priority- so the likelihood of being sued will weigh more heavily then the probable harms to the individuals concerned. This calculus completely avoids the test that a good person will take good decisions and consider individuals from a different perspective. The organizational milieu ensures that this conflict will rarely if ever be called to account in any other way than a legal risk.

The ethical issues raised by the use of Big Data range from collectors, assemblers, analysts and exploiters of the results. In itself this is a normal segmentation of the ethical issues and responsibilities, but the ethical issues of each segment do not necessarily reflect the overall ethical concerns that are raised. The aggregators might perhaps be seen to be those at greatest ethical risk, as diverse sets of data can now be used to de-anonymise most sources of personal or related data. However the moral values of the other layers are equally important.

Good business decisions may not be good ethical decisions (Hartmann, 2013) Hartmann gives a clear example of a choice for the individual concerned as to which to choose. A virtue ethics stance would incline towards the ethical rather than the more convenient (and more easily rationalized) utilitarian harms v benefits standpoint by the conflicted person): rationalizing which is more beneficial in societal terms is in this case simply a way of avoiding a clear ethical decision based on a good character: it also raises questions of personal v organisational trust. Which is more dispensable? To the individuals concerned?

The harms balance that a utilitarian analyst would be primarily concerned

with are also diffused down this chain of different Big Data functions and responsibilities. But the moral values (virtue ethical stances) of people involved at each level are considerably more difficult to assess than in Hartmann's examples. This applies equally to the people themselves in their own situations, and to philosophical and ethical analysts observing the overall process.

One line of argument is to consider the full chain to be an organisational responsibility - which immediately places the moral choices of management in the firing line - although they might well be completely ignorant of the ethical issues (or possible transgressions) made as a consequence of their broad level instructions to the marketing or other departments who undertake the big data and analytics task and operational decisions.

This point is approached but skirted around and missed (Zwitter, 2014), while acknowledging the power imbalances that arise from Big Data strategies (usually to the benefit of well resourced corporate entities). However Zwitter correctly identifies the key point of our argument on sliced and diced diffusion of responsibility due to the complexity and diversity of stages in Big Data utilization as follows:

"it seems not to be an overstatement to say that Big Data does have strong effects on assumptions about individual responsibility and power distributions" (Zwitter, 2014)

This places the onus on our arguments in this paper exploring the utility of a Virtue Ethics stance, rather than a harms v benefits utilitarian consequential

impacts balance.

Unfortunately, Zwitter did not recognize this distinction. Extending his arguments does lead to this. Even in areas where moral issues such as the Right to be Forgotten are recognized, the public/academic debate leans heavily on the harms v benefit framework, and not on any intrinsic valuation of the humanity issues - or indeed the moral worth or otherwise of a decision once made, but rather relies on careful framing of law (which of course in most cases none could possibly afford to use to contest: a common problem when legal frameworks alone are relied upon, on the theory that those without resources(usually the most vulnerable) have financial unaided access to the courts to determine previously untested laws). A typical assertion is as follows:

> ' *Dynamics of Privacy and Publicity, refined by differentiating the condition for an injunction and for a sanction (compensation and fine. The injunction should be conditioned to the mere reversal of the balance of legal interests, while the sanction should require the unreasonableness of the assumption that the reversal has not yet taken pace, or the moral certainty that a reasonable decision-maker would assume that the reversal has taken place. With regard to the regulation of the right to be forgotten, I would argue, a combination of the latter approaches is needed, possibly with a preference for the third approach with regard to the uploader and for the second approach with regard to the provider'* (Sartor, 2014)

It is difficult to argue that corporate ethics should not accept a personal moral stance when the application of Big Data based strategies and

segmentations are probabilistic yet still have substantial potential impacts on individuals as if it were a factually identified allocations to specific groups for differential treatment (a form of automated stereotyping). This aspect, of group misallocations, can be characterised by the following process:

- The marshaling of the data analytics is to determine probabilities that lead to actions
- The actions are therefore founded on almost certainly 'faulty' and unverified (and, due to scale and non-commensurate data structures, unverifiable) consequently accountability cannot therefore be relied upon creating moral hazard for management
- The outcomes of big data conclusions are therefore from the Intelligence background and values and culture.
- They are certainly not consistent with civil law, chains of evidence and verifiability

This raises a wide range of ethical issues from both utilitarian and virtue ethics standpoints: one from the risk assessment judgments that underpin utilitarian ethical stances, and the moral worth frameworks for those responsible: a virtue ethics perspective.

The managerial perspective can summarised nicely as balance between Agency theory (opportunistic self interest) and Governance/Stewardship (organisational verifiable accountability processes) (Brooks & Dunn, 2012). Neither gives much weight to the (personal) moral hazard encountered by managers.

These are already raising unresolved (and possibly irresolvable) ethical problems for and within organisations.

One of the escape routes is to rely upon are formal processes. These reduce the reliance on individual responsibility for difficult decisions, but are available only where reliable due process legal frameworks are in place. Australia is hardly very distinguished in its adoption, use and support for such protections:

> *'ICT companies should wield their great power responsibly, which means adhering to the principles of digital due process that are implicit in the foundational instruments of the International Covenant of Civil and Political Rights, the European Convention on Human Rights, and the United States Constitution*' (Nunziato, 2014)

Big Data can be considered as a domain of Information Ethics (Ess, 2009), but strong arguments are raised by Doyle (Doyle, 2010) that this stance is insufficient to justify a non-utilitarian perspective. However we concur that:

"*What we are discovering is that we need an augmented ethics for a theory of augmented moral agency*" (Floridi, 2008).

Values are often well articulated by religious leaders. The Second Encyclical of Pope Francis places power, ethics, technology and morality- ineffably the issues at the heart of Big Data Ethics- in context

'The fact is that "contemporary man has not been trained

to use power well" (Guardini, 1998) because our immense technological development has not been accompanied by a development in human responsibility, values and conscience' (Pope Francis, 2015)

It would be difficult to find a more apposite skewering of the modern habit to implement institutional ethics by formal processes, and a studied layering of protective differentiated responsibilities that diffuse the central issue: ethics and moral decisions are taken by people, not by organisations.

The EU Data Protection Commissioner is in general accord with this perspective, with structured procedural recommendations as follows, closing with the necessity of 'empowered individuals':

> *'a four-tier 'big data protection ecosystem' to respond to the digital challenge: a collective effort, underpinned by ethical considerations.*
> *(1) Future-oriented regulation of data processing and respect for the rights to privacy and to data protection.*
> *(2) Accountable controllers who determine personal information processing.*
> *(3) Privacy conscious engineering and design of data processing products and services*[5]*.*
> *(4) Empowered individuals.* '(European Data Protection Supervisor, 2015)

One of the major aspects of Big Data utilisation is the progressive use disintermediated management, alterations to governance driven by data and

[5] An area found to be fitted best as *egocentric consequentialism*, ie *as a process of social negotiation rather than as technological progress* (Lloyd & Busby, 2003)

consequent decisions flows, and changes to some of the assumptions - and indeed power structures - on which current organizational conventions are founded. An excellent paper (Rosenblat & Stark, 2015) analyses the functioning of Uber, where the processes of management control and management directives are systematically disintermediated to an algorithm. We will see more and more of this growing process of management disintermediation.

This is a practical illustration of the points made about Smart Cities, where massively enhanced data flows and sensors are predictably leading to exactly this kind of still largely unrecognized governance stressors (Wigan, 2015).

Ethical responses to these pressures and changes press one to take the Virtue Ethics stance far more seriously, where parties who do retain decision and decision capacity are held to personal ethical standards, rather than a simplistic utilitarian perspective. Business ethics will always present conflicts between employer and employee moralities, and the quantum leap of data flows in quantity (and unverifiability) will soon make these questions ones for a steadily broader range of people and organisations.

Conclusions

Like most areas of globalised mass information the legal domains are unclear and inconsistent, and the ethical problems have no consistent basis – other than the Virtue Ethics stance of the person doing the work or apply it.

Utilitarian approaches have no such problem, which would seem to be good grounds for complementing utilitarian calculus for clarification of personal ethical stances, where virtue ethics has clear guidance to give. In such an alternative framework, some of the tradeoffs (harm minimisation) implicit in utilitarian approaches will need to be complemented with areas where some tradeoffs are quarantined.

We contend that a mix of both utilitarian and virtue ethical perspective is essential to address the complexity and multilayered nature of the issues raised by the collection, aggregation, analytics and utilisation of Big Data by organisations.

References

Anon. (2015). *2015 Black Hat Attendee Survey: Time to Rethink Enterprise IT Security*. Retrieved from https://www.blackhat.com/docs/us-15/2015-Black-Hat-Attendee-Survey.pdf

Aspers, P. (2007). Nietzsche's Sociology. *Sociological Forum, 22*(4), 474-499.

Beaulac, S. (2004). The Westphalian model in defining International law: challenging the myth. *Australian Journal of Legal History, 8*. Retrieved from http://www.austlii.edu.au/au/journals/AJLH/2004/9.html

Boos, M. (2003). *International transfer pricing the valuation of intangible assets*. The Hague: Kluwer International.

Branscomb, A. W. (1994). *Who owns information? From privacy to public access*. New York: Basic Books : Harper Collins.

Brooks, L. J., & Dunn, P. (2012). *Business and Professional Ethics for Directors, Executives, and Accountants* (6 ed.). Mason Ohio: Cengage.

Caporaso, J. A. (2000). Changes in the Westphalian Order: Territory, Public Authority, and Sovereignty. *International Studies Review,, 2*(2), 1-28.

Contractor, F. J. (Ed.) (2001). *Valuation of intangible assets in global operations*: Greenwood Publishing.

Corrado, C., Hulten , C.,.and Sichel, D. (2005), Measuring capital and technology: an expanded framework. (2005). Measuring capital and technology: an expanded framework. In J. H. a. D. S. C. Corrado (Ed.), *Measuring capital in the new economy* (pp. 65, 11-45). Chicago: Chicago University Press.

Daigle, C. (2006). Nietzsche: Virtue Ethics ... Virtue Politics? *Journal of Nietzsche Studies*(32), 1-21.

Donelley, R., Thompson, R. G., & Wigan, M. R. (2012). Process validation of urban freight and logistics models. *Procedia Social and Behavioural Sciences, 39*, 400-408. doi:http://dx.doi.org/10.1016/j.sbspro.2012.03.117

Doyle, T. (2010). A Critique of Information Ethics. *Knowldege Technology and Policy, 23*, 163-175. doi:10.1007/s12130-010-9104-x

Ess, C. (2009). Floridi's Philosophy of Information and Information Ethics: Current Perspectives, Future Directions. *The Information Society, 25*(3), 159-168. doi:10.1080/01972240902848708

European Data Protection Supervisor. (2015). *Opinion on 4/015: Towards a new digital ethics: Data, dignity and technology*. Retrieved from

Floridi, L. (2008). Information ethics: a reappraisal. *Ethics and Information Technology, 10*(2-3), 189-204.

Floridi, L. (2015). The anti-counterfeiting trade agreement: the ethical analysis of a failure, and its lessons. *Ethics of Information Technology*. doi:10.1007/s10676-015-9374-9

Guardini, R. (1998). *'The End of the Modern World' and 'Power and Responsibility' combined single volume*. Wilmington USA: Isi Books.

Hales, S. D. (2000). Recent Work on Nietzsche. *American Philosophical Quarterly, 37*(4), 313-333. Retrieved from http://www.jstor.org/stable/20010008

Hartmann, E. (2013). The virtue approach to business ethics. In D. C. Russell (Ed.), *The Cambridge Companion to Virtue Ethics* (pp. 240-264). Cambridge: CUP.

IEEE Standards Association. (InProgress). P2413 - Standard for an Architectural Framework for the Internet of Things (IoT). http://standards.ieee.org/develop/project/2413.html: IEEE.

Kehly, D., Bankston, K., Greene, R., & Morgus, R. (2014). *Surveillance Costs: The NSA's Impact on the Economy, Internet Freedom & Cybersecurity*. Retrieved from Washington DC: https://www.newamerica.org/downloads/Surveilance_Costs_Final.pdf

Kushner, D. (2013, 26 Feb). The Real Story of Stuxnet: How Kapersky tracked down the malware that stymied Iran's nuclear-fuel enrichment program. *IEEE Specturm.* Retrieved from http://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet

Lloyd, P., & Busby, J. (2003). "Things that Went Well - No Serious Injuries or Deaths": Ethical reasoning in a Normal Engineering Design Process. *Science and Engineering Ethics, 9*, 503-516.

Martin, C. W. (2006). Nietzsche's Virtues and the Virtues of Business. In J. Welchman (Ed.), *The Practice of Virtue: Classic and Contemporary Readings in Virtue Ethics* (pp. 204-224). indianapolis: Hackett.

Mathews, A., & Tucker, C. (2014). Government Surveillance and Internet Search Behavior. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2412564

Mellody, M. R. (2014). *Training students to extract data from Big Data: Summary of a Workshop*. Washington DC: National Academies Press.

Monga, V. (2014, October 12). The Big Mystery: What is Big Data Really Worth? A Lack of Standards for Valuing Information Confounds Accountants, Economists. *The Wall Street Journal*.

Nunziato, D. C. (2014). With Great Power Comes Great Responsibility: Proposed Principles of Digital Due Process for ICT Companies. In L. Floridi (Ed.),

Protection of Information and the Right to Privacy - A New Equilibrium? (pp. 63-84): Springer.
Oakley, J., & Cocking, D. (2001). *Virtue Ethics and Professional Roles*: Cambridge University Press.
Pope Francis. (2015). *Encyclical letter "Laudato Si" of the Holy Father Francis on Care for our common home*. Rome Italy: Vatican Press.
Rosenblat, A., & Stark, L. (2015). *Uber's Drivers: Information Asymmetries and Control in Dynamic Work*. Paper presented at the Winter School: Labour in the on-demand economy, Centre for European Policy Studies. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2686227
Russell, D. C. (2009). *Practical Intelligence and the Virtues*. Oxoford: Clarendon Press.
Sartor, G. (2014). The Right to be Forgotten: Dynamics of Privacy and Publicity. In L. Floridi (Ed.), *Protection of Information and the Right to Provacty- A New Equilibrium* (pp. 1-15): Springer.
Schneider, K. F., Lyle, D. S., & Murphy, F. X. (2015). Forum: Rethinking the Cyber Domain and Deterrrence: Framing the Big Data Ethics Debate for the Military. *JFQ*(77).
Swanton, C. (2003). *Virtue Ethics: A pluralistic view*. Oxford: OUP.
Teilhard de Chardin, P., & Tr: Bernard Wall. (1959). *The Phenomenon of Man*. London: Collins.
Teilhard de Chardin, P., & Tr: Norman Denny. (1964). *The Future of Man*. London: Collins.
Wigan, M. R. (1996). The problems of success: privacy, property, and transactions. In L. Branscomb & J. Keller (Eds.), *Converging Infrastructures: Intelligent Transportation and the NII* (Vol. 1, pp. 341-354). Cambridge, Mass.: MIT Press.
Wigan, M. R. (2010). Owning identity - one or many - do we have a choice? *IEEE Technology and Society Magazine, 29*(2 (Summer)), 33-40. doi:http://dx.doi.org/10.1109/mts.2010.937026
Wigan, M. R. (2015). Smart Cities: Environmental Aspects. *IEEE Smart Cites Launch Workshop.* Retrieved from http://works.bepress.com/mwigan/17
Zwitter, A. (2014). Big Data Ethics. *Big Data & Society, 1*(2), 1-6. doi:10.1177/2053951714559253