Emerging Concerns for Responsible Data Analytics: Trust, Fairness, Transparency and Discrimination

Paper for the NSW Data Analytics Centre Showcase, 12 July 2017

Peter Leonard[1]
Principal, Data Synergies
Advisory Board Member, NSW Data Analytics Centre

The privacy assessment framework governing uses of information about individuals is now mature, understood and applied by responsible data custodians. However, the privacy assessment framework is incomplete. The privacy framework focusses assessment of data analytics projects upon responsible management of personal information about individuals as a key consideration. But this is not the only important consideration. This narrow focus is now diverting attention from emerging concerns as to maintenance of consumer trust, fairness of outcomes and adverse impacts of uses of insights derived from data analytics. Building torrents of academic papers elucidate concerns as to trust, ethics and fairness. Often these concerns are grouped together by academic commentators under a rubric of 'data ethics' and coupled with calls for responsible data custodians to apply ethical principles to identify and address 'unethical' outcomes of data analytics projects. Discussions by academic commentators canvass concerns as to social equity, corporate social responsibility, expectations of 'transparency', concerns as to 'unaccountable algorithms', the philosophy of ethics and the weighting of benefits for many against detriments for a few. But this discussion does not provide a practical process to address these concerns in the course of project management processes as currently used in businesses and by government agencies.

Some expert commentators, notably including Marty Abrams and Peter Cullen at the Information Accountability Foundation, have suggested a framework integrating ethical evaluation within a privacy impact assessment. This paper suggests that such integration is not practicable and in fact sub-optimal, although coordinated assessment of fairness and ethics is practical and often necessary. This paper also suggests that it will often be appropriate to ensure that responsible assessment is conducted as to uses and applications of outputs (such as algorithms or insights) of data analytics projects in circumstances where these outputs do not themselves constitute uses or disclosures of personal information that are (or should be) subject to privacy assessment. Algorithms may go into productive use in circumstances where limitations as to underlying data used to generate such algorithms are not understood and their reliability is affected by exogenous factors, where the algorithms is inherently biased, or where a particular application of the algorithm has unfair discriminatory effect.

The objective of this paper is to promote development of a framework for phased assessment of data analytics projects that encompasses ethical and fairness considerations while not creating a dead weight of multiple detailed assessments. The framework must be targeted, agile and capable of application by responsible teams that are not ethicists. These teams should be empowered to ask and seek answers to sensible questions framed in plain language. This paper endeavours to ensure that structured questions are asked at the right phase of a data project.

[1] Peter Leonard is a data and technology business consultant and lawyer and principal of Data Synergies, a new data commercialisation consultancy. Peter was a founding partner of Gilbert + Tobin. Following his retirement as a partner in 2017 he continues to assist Gilbert + Tobin as a consultant. Peter also chairs the Australian IoT (Internet of Things) Alliance's Data Access, Use and Privacy work stream and the Law Society of New South Wales Privacy and Communications Committee. The IoT Alliance is Australia's peak body bringing together industry and governments to address issues affecting IoT adoption and implementation. He also participates in the Australian Computer Society's Data Taskforce as chaired by Dr Ian Oppermann, NSW Chief Data Scientist. Peter wishes to acknowledge the contribution of all Taskforce members to taxing discussions within the Taskforce that have included exploration of concepts developed in this paper.

1 Introduction and key points

Data linkage of data sets between organisations is a young, but fast growing, area of data science practice. Trusted third party arrangements are a key aspect of controlled data sharing that address many of the privacy concerns that arise in relation to data release into the public domain. However, maintenance of trust of consumers and citizens as to what they understand to be "data sharing" depends upon businesses and governments engaging with concerns of many citizens about privacy and security and also undertaking responsible assessment as to fairness and equity in use of outputs of data sharing. There are legitimate public concerns about unconstrained data sharing. In an environment of declining public trust in business and government[2], citizens and their advocates should not be expected to rely upon assurances made by governments and businesses as data custodians and data sharers that they can be trusted to meet community expectations.[3]

This paper considers the circumstances in which data analytics projects require more than privacy impact assessment. Individuals may suffer adverse effects from otherwise properly privacy managed projects because these effects flow from implementation of outcomes which do not involve any relevant use or disclosure of personal information. There is a clear gap between privacy impact assessment and human research ethics review. Properly privacy managed data projects outside the fields of human research[4] may not be subject to any form of ethics review or assessment. Review by human research ethics committees (**HRECs**) of data projects is notoriously cumbersome and slow.[5] Sometimes HREC reviews are not fully informed by understanding of issues that may now arise through algorithmic decision making and applications of machine learning that may adversely affect some individuals. Accordingly, HREC review processes are not be a suitable model for a broader and more agile ethics assessment of data projects that are not human research projects. Moreover, privacy impact assessments (**PIAs**) are often conducted before the discovery phase of a complex data analytics project, at a time where outputs and outcomes from application of those outputs cannot reasonably be anticipated and assessed.

It is time to fill the gap between project initiation PIAs and any application phase consideration of benefits and adverse effects of then specified outcomes. The solution requires revisiting a PIA after the discovery phase, to check that any analysis of re-identification risks associated with outputs remains current and appropriate. The solution also requires a more fundamental rethinking of scope and utility of PIAs. In rethinking that scope and role, it is suggested that existing human research ethics review processes are too cumbersome, slow and complex for non-medical data analytics projects. A different approach to outcomes assessment is now required.

Project outcome assessments often will not be possible at the discovery phase of a project. Provided that data inputs and working data are properly privacy protected, adverse effects should only occur if outcomes are generated from the application phase of a project. Often the various possible outcomes can be properly identified and scoped late in the conduct of the application phase of a data analytics project. Where there are reasonably foreseeable unfair or otherwise adverse effects of possible project outcomes upon individuals, the degree of effect upon individuals should be considered in a well-executed project outputs assessment conducted early in the application phase of any significant data analytics project. This project outputs assessment should be conducted wherever adverse outcomes upon some individuals are reasonably foreseeable, regardless of perceived social benefits for many and other perceived social utility of a project, and regardless of whether personal information is being used in any part of the project.

---

[2] See for example *The Edelman Trust Barometer 2017* at http://www.edelman.com/trust2017/. As there stated, "The 2017 Edelman Trust Barometer reveals that trust is in crisis around the world. The general population's trust in all four key institutions - business, government, NGOs, and media - has declined broadly, a phenomenon not reported since Edelman began tracking trust among this segment in 2012. With the fall of trust, the majority of respondents now lack full belief that the overall system is working for them. In this climate, people's societal and economic concerns, including globalization, the pace of innovation and eroding social values, turn into fears, spurring the rise of populist actions now playing out in several Western-style democracies. To rebuild trust and restore faith in the system, institutions must step outside of their traditional roles and work toward a new, more integrated operating model that puts people - and the addressing of their fears - at the center of everything they do."

[3] Many media reports have chronicled the string of privacy related concerns arising as to uses of data by Australian government agencies over the 2016-17 Australian financial year. By way of examples, see articles in The Mandarin including Matthew Beard, 6 March 2017, '*When it comes to trust, a good offence is your worst defence*' http://www.themandarin.com.au/76454-high-price-to-pay-to-correct-the-public-record/ and Anna Johnston, 30 June 2017, '*A litany of privacy disasters: how to ruin public faith in just 12 months*' http://www.themandarin.com.au/80791-litany-privacy-disasters-ruin-public-faith-just-12-months/.

[4] See further Office of the Australian Information Commissioner, *Health information and medical research* https://www.oaic.gov.au/privacy-law/privacy-act/health-and-medical-research; Australian Government National Health and Medical Research Council, *Human Research Ethics Committees (HRECs)*, https://www.nhmrc.gov.au/health-ethics/human-research-ethics-committees-hrecs.

[5] See for example the discussion in Productivity Commission *Data Availability and Use Inquiry Report,* No. 82, 31 March 2017, particularly at pp 509-540.

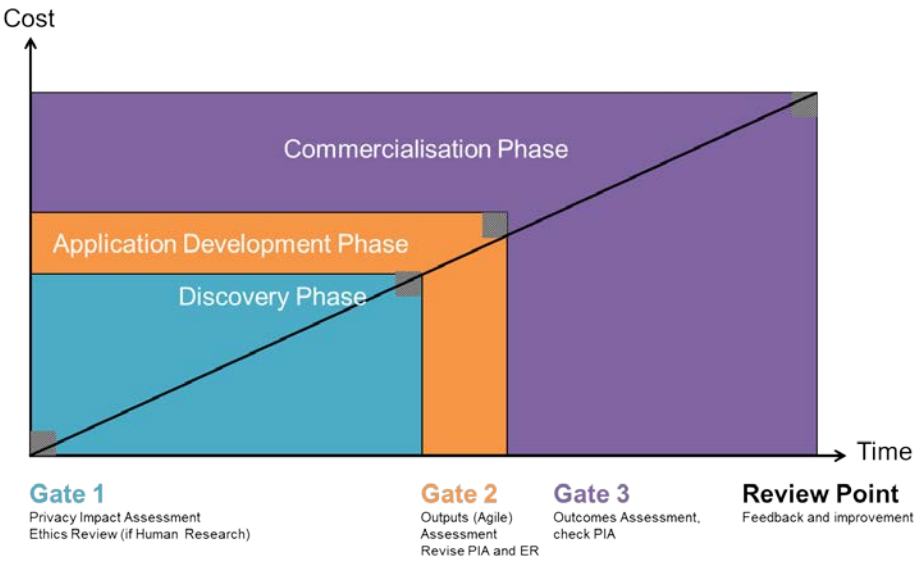
A project outputs assessment should then be revisited by conduct of an outcomes assessment when the applications are developed and the outputs (be they insights, algorithms or processes or methodologies) are ready for application. The outcomes assessment often will be a straightforward review to confirm that the outputs assessment was sufficiently scoped to anticipate and address possible outcomes arising through application of the outputs. However, and sometimes, the outcomes assessment process may lead to recommendations as to caveats or qualifications as to uses of outputs that should be notified to prospective users of those outputs, or as to publication to ensure appropriate transparency and accountability as to application of outputs (to address otherwise reasonably anticipated public concerns as to possible adverse applications and outcomes).

Processes for outputs assessment and outcomes assessment must walk a fine line between being unduly cumbersome and slow and 'box ticking' superficiality. Assessment processes must not be 'window-dressing' or formulaic application of cost benefit analysis undertaken to provide defence or self-justification for a government or business unwisely planning to undertaking a particular court of activity.[6] Assessment must be fair and balanced, or trust in such processes will rapidly erode.

Sometimes applications of outcomes lead to unexpected results or provide other feedback that should inform further uses of those applications. So the outcomes assessment should be revisited after a period in operation, to see how the application should be refined or otherwise changed. This paper suggests that a review assessment should be conducted, say, at 12 months out from first operational use.

The recommendations in this paper may be graphically represented, as follows:

Figure One: Data Analytics Project Review Framework



Of course, conduct of a PIA should be (and remain) the first step in any properly planned data analytics project of scale that uses personal information about individuals as an input.[7] In the context of privacy protected data linkage (as discussed below in this paper), the PIA would ensure that relevant data linkage is conducted in a properly

[6] The UK Cabinet Office's *Data Science Ethical Framework*, ver. 1.0, May 2016 (available at https://www.gov.uk/government/publications/data-science-ethical-framework) has been stringently criticised by some commentators as being oriented towards self-justification rather than more balanced impacts assessment. See for example Roger Clarke and Charles Raab, 'Inadequacies in the UK's Data Science Ethical Framework', *Euro. Data Protection L. 2, 4* (Dec 2016) 555-560, also at http://www.rogerclarke.com/DV/DSEFR.html. Foe examples of such formulaic assessments, see many *Statements of Compatibility with Human Rights Prepared in accordance with Part 3 of the Human Rights (Parliamentary Scrutiny) Act 2011* in Explanatory Memoranda that accompany Bills introduced into the Australian Federal Parliament.

[7] See OAIC, *Privacy management framework*, May 2015 as available at http://www.oaic.gov.au/privacy/privacy-resources/privacy-guides/privacy-management-framework; also *Guide to developing an APP privacy policy* and *Guide to undertaking privacy impact assessments,* each May 2014 and available at http://www.oaic.gov.au/images/documents/privacy/privacy-resources/privacy-guides/guide-to-developing-an-app-privacy-policy.pdf and http://www.oaic.gov.au/images/documents/privacy/privacy-resources/privacy-guides/guide-to-undertaking-pias.pdf.

controlled and safeguarded data eco-system, such that linked information may be considered secure and properly de-identified, albeit not fully anonymised, and all outputs are risk assessed as low or remote re-identification risk.[8]

Accordingly, this paper now addresses states that first step.

2 Concerns with data linkage projects t

Data linkage projects typically require accommodation of concerns of a variety of stakeholders. Concerns typically addressed include:

- where personally identifying information is to be disclosed for and used as personal information in data linkage, that the use is for a purpose of which affected individuals had notice or, in the case of personal information that is sensitive information, for which affected individuals provided fully informed and voluntary consent,
- whether the data custodian is able to reliably verify that affected individuals had that notice or provided that consent,
- information security, including guarding against threats from internal unauthorised intrusion and external threats including malicious attacks (denial of service, hacking etc.) and cyber-espionage,
- ensuring clarity as to who 'owns', maintains and is responsible for control of distribution of data (for example as to which core data sets may be transformed by cleansing, normalising, key coding, merged or other transformations or value adds, as to ownership and of subsequent use of transformational code, algorithms and inferences, insights, and reports derived from data analytics conducted on these data sets),
- maintaining trust of citizens that information about them will not be used by government agencies or business enterprises in ways that are privacy invasive, 'spooky', contrary to accepted societal norms from time to time, or in ways that may lead to them suffering unfair adverse consequences,
- complying with restrictions in contracts and in statutes, and
- protecting business confidential information and trade secrets.

Many data sharing projects fail to proceed due to an inadequate framework to resolve privacy or 'trust' concerns.

3 Privacy risk management: de-identification frameworks[9]

To date privacy concerns around data sharing have often been addressed by using 'masking' of identifiers: that is, the removal of personal identifiers and the pseudonymisation of data sets using transactor keys or tokens. However, some privacy advocates assert that technological advances and multiplicity of data points make re-identification of individuals from pseudonymised data relatively straightforward. In response, there has been extensive work in recent years in developing privacy protective risk management methodologies in order to specify appropriate and legally enforceable requirements for data linkage of data about individuals. These methodologies may be employed to properly protect data sets such as card transaction records, geo-located movement traces and patient level epidemiological health data, that otherwise may be vulnerable under reidentification attack.

Privacy risk management turns on recognising a distinction between de-identification and anonymisation. The stages of removal or obfuscation of direct (name) or indirect (mobile number, movement trace, email address) identifiers of any individual included in a stream of transaction data can be seen as steps along a continuum of de-

[8] There is a rapidly growing literature on good privacy and security practice in anonymisation and de-identification. See in particular UK ICO, *Anonymisation Code of Practice* and associated material at https://ico.org.uk/for-organisations/guide-to-data-protection/anonymisation/; the UK Anonymisation Network, *Anonymisation Decision Making Framework* and associated materials at http://ukanon.net/ukan-resources/ukan-decision-making-framework/; U.S. National Institute of Standards and Technology (NIST) *De-identifying Government Datasets*, NIST Special Publication 800-188 (Second Draft), *De-Identification of Personal Information*, NISTR 8053, Information and Privacy Commissioner of Ontario, *De-identification Guidelines for Structured Data*, June 2016, Article 29 Data Protection Working Party, *Opinion 05/2014 on Anonymisation Techniques*, Omer Tene, Kelsey Finch and Jules Polonetsky '*Shades of Grey: Seeing the Full Spectrum of Practical Data De-Identification*', Santa Clara Law Review, forthcoming; UK ICO, *Big data, artificial intelligence, machine learning and data protection*, March 2017.

[9]

identification. Effective anonymisation of transaction level information is the logical end point of that continuum. Anonymisation means the data transaction information still addresses a unique and distinct transactor, but does not enable the individual that is the unique transactor to be identified, whether from the information itself or from any combination of data points reasonably available to any entity that has access to the data stream or its derivations.

Of course, de-identification to the point of anonymisation can often be achieved by aggregation of individual data points, typically for the purpose of making comparisons or identifying patterns: that is, to show general trends or values without leaving granular indirect identifiers that might leave an individual identifiable within the data. Applying *k*-anonymity or like methodologies, values determined to be of 'small numbers' may be supressed to minimise risk of reidentification, either through blurring or through omission altogether. Sometimes it is possible to de-identify data to the point where the transformed data is safe for public release because there is no more than a remote risk of individuals being identified: in this case, the data has been effectively anonymised. Of course, once the data is released the full artillery of re-identification techniques may be employed on the data by anyone, so anonymisation must be particularly robust, including over time.

Unfortunately, the utility of effectively anonymised data for many purposes, and particularly for epidemiological applications, is severely compromised by aggregation, suppression or blurring. In such cases, alternative measures must be taken that retain the usefulness of unique individualised data whilst still protecting the privacy of the individuals concerned. Clearly, useful individual level data should not be released publicly – the risk of re-identification of at least some individuals would be very high - but re-identification risk associated with data analytics uses may be managed through pseudonymisation combined with controls as to access and application of that data. This scenario may be referred to as controlled (or safeguarded) release, only for use in a recognised 'de-identification zone'. In the scenario of controlled release, the assessed harms from re-identification may be allowed to be higher than for data released into the wild. Assessed risk is a measure of the extent of threat by a potential circumstance or event and so typically a function of both the adverse impacts that would arise if the circumstance or event occurs, and the likelihood of occurrence. Therefore, controls deployed in the safeguarded data environment may substantially reduce likelihood of attempts at re-identification while in that environment. Physical, system, human and permitted output controls as the perimeter of the safeguarded de-identification zone may ensure that outputs from that environment are appropriately aggregated or otherwise privacy protected. Thus, assessed harms from improper release of particular data sets from the safeguarded data environment may be high, but the risk that those harms will be suffered may be so effectively mitigated that assessed re-identification risk is remote.

In summary, data sets and data streams that would usually be considered too high risk to individual privacy protection may be managed within a properly planned, documented and implemented privacy management framework that can reduce re-identification risk to the point where this risk is remote within the particular context of controlled release and use. A similar risk assessment methodology may be applied to both controlled access and public release data sets, to determine the point at which re-identification risk is sufficiently remote for the particular context of use.

4 Privacy risk assessment and risk mitigation

Two important consequences follow from assessment of risk requiring both assessment of harms and likelihood of those harms being suffered.

First, where data is de-identified for limited disclosure or access, provided that disclosure or access has been appropriately, reliably and verifiably limited and controlled, re-identification risk will be significantly less than if that same data was put into the wild.

Second, likelihood of occurrence might be mathematically expressed using an objective scale. But because harms are likely to be quite specific to the circumstance and particular individuals, a fact specific, contextual analysis is required. In any event, there is no regulatory clarity as to how a 'low' or 'remote' point of likelihood of occurrence is to be objectively and statistically measured, so assessment of risk has an inherently subjective element. That is why privacy impact analysis remains an inexact science, or as some information security experts see it, something of a black art. In any event, an information privacy and fairness assessment should be conducted in relation to any project involving use of purportedly de-identified data which carries any reasonably ascertainable risk of re-identification of any affected individuals, both as to risks within the safeguarded data environment and as to outputs from that zone which themselves might be personally identifying.

These are areas where contractual restrictions imposed on data projects at their inception are particularly important, along with specification of processes and procedures to ensure that these contractual restrictions are understood, followed and verifiably reliable.

Often data linkage projects are outsourced to third parties, to leverage their data science skills and methodologies and to create separation from a data custodian's personally identifying data sets. Data analytics services providers may in controlled environments facilitate privacy protective data linkage of individual level data. Relevant controls vary, but privacy and security by design compliant arrangements for data linkage of individual level data about individuals are typically based upon **five key control elements**:

- **separation** of persons or entities with access to personally identifying information from those persons or entities ('trusted third parties') conducting analytics using data sets which have been pseudonymised;

- replacement of direct or indirect personal identifiers in the merged data sets with a linkage code, or **transactor key**, which enables the service provider to infer that an identifiable transactor found in each data set is a unique transactor, although not identifiable;

- a combination of technical, operational, contractual and otherwise legally enforceable **safeguards** which reliably and verifiably ensure that uses of data outputs are only in accordance with stated purposes and that individuals that are the subject of transaction data are not re-identified and that records of personal information about those individuals held by any relevant party are not augmented or supplemented in any way through the controlled process,

- **information governance** oversight, data process controls, change control procedures and quality assurance processes that ensure that each of these things are reliably and verifiably implemented and then reliable in ongoing operation and that any change in data flows or deviation from required practices and procedures is promptly identified, considered and (if need be) addressed by appropriate risk mitigation measures.

- **output controls.** All the above will be useless if insights released out of the controlled environment are not appropriately privacy protective.

Such arrangements are sometimes called **trusted third party arrangements**. However, the requirements that engender and enable 'trust' should be embodied in specific contractual obligations and associated work processes and procedures to ensure that the arrangements are appropriately privacy protective. These arrangements are accordingly not a matter of 'trust'. The requirements are legally enforceable - and often exacting to meet and to verify.

Of course, exacting de-identification requirements are not required if affected individuals have given fully informed consent to particular data sharing. Consent is always the best solution, but often this solution is not available because a proposed release and use was outside the contemplation of the data custodians when consents were obtained and the relevant data collected. So often a view will need to be taken as to whether the act or practice of creating and using data linkage code is an act or practice of a data collector regulated by privacy law. Legal questions then arise as to:

- Whether pseudonymisation of personal identifiers itself an act or practice in relation to personal information which requires notice to or consent of the affected individual (and if so, how express that notice or consent needs to be), or is pseudonymisation an act or practice is akin to, say, anonymisation of personal information by aggregation up in reports and analyses?

- To what extent can a party disclosing deidentified transaction level information and associated data linkage code rely upon that party's assessment as to the likelihood of compliance of a downstream recipient with relevant prescribed safeguards? In other words, how active must the discloser be in verifying that a downstream recipient will meet such commitments as the recipient is willing to give as to its compliance with the requirements which underpinned the discloser's decision to facilitate the data linkage?

The answers to these questions in the Australian regulatory environment remain the subject of debate and disagreement.

5 Looking downstream: reasonableness and fairness

We have already noted that perimeter controls around a controlled environment may ensure that released outputs from that environment are appropriately aggregated or otherwise privacy protected. This aspect of safeguards requires particularly careful attention. As is now widely recognised, it is important to ensure that permitted inferences, insights and reports derived from data analytics conducted on the safeguarded data sets do not leave any individual reasonably identifiable. By contrast, some data analysts neglect to ensure:

- **ongoing protection of source data**: that the source (underlying) data remains protected in accordance with expectations and requirements of the contributor data custodians;
- **reasonableness**: that uses and applications of inferences and insights remain properly grounded in subsequent derivations and re-presentations through appropriate statement as to limitations of the source data and any possible exogenous factors;
- **fairness**: that uses and applications of inferences and insights are fair.

A good example of the problem is afforded by the Facebook emotional contagion study published in 2014[10]. In this paper, Facebook data scientist Adam Kramer and Cornell University social scientists Jamie Guillory and Jeff Hancock reported as to the results of experimentally modifying the Facebook feed algorithm for 689,003 people. They demonstrated that the negative or positive emotional valence of the posts that show up on users' News Feed alters the emotional valence of the posts that the user him or herself makes. This supported the hypothesis that emotional contagion—spreading emotional states through social contact—occurs on a massive scale, albeit with relatively small individual effects, in social networks.

This paper launched a public controversy about big data research ethics, principally because Facebook users whose feeds were altered did not directly consent to participating (although it was suggested that they did so somewhat obliquely by inferred consent to Facebook's then terms of services). Some critics questioned the ethics of Internet services experimenting with their users' emotional states with low standards as to consent. Other commentators suggested that the major difference between this study and common online business practices was that that the results of this experiment were made public and then through publication in a scientific journal. In any event, the study had not been subjected to any form of ethics review, and probably was not required to be so reviewed because it was not an 'intervention' in the life of human subjects by way of '(human) research'. Some critics suggested that the problem went deeper, suggesting that the data science community was not familiar with the informal modes of ethics regulation found in other science and technology communities. The observation was made that the three core disciplines that inform the nascent field of data science (computer science, physics, and applied mathematics) have long been considered as outside of human-subjects-related ethics concerns because their work and contributions have historically been about systems and not people. As a result, it was suggested that data scientists are not yet familiar with assessment and management of ethics concerns.

Whatever the reason, the Facebook emotional contagion study illustrates dangers of data analytics projects which are not privacy invasive in a traditional sense and not subject to any form of fairness/ethics/trust review.

In the balance of this paper we further explore how this fairness/ethics/trust review might be built into data projects.

6 Other non-privacy related effects or outcomes

Increasingly, we need to anticipate concerns as to **non-privacy related unfair effects or outcomes** and to then seek to address them. At this point, we move beyond the current requirements of privacy law and into more indeterminate, non-legal concerns. The fact that these concerns are more indeterminate does not make addressing them any less important. However, contention as to both the nature and extent of these concerns means that they are often underrated or simply dismissed. Are we talking about:

[10] Kramer ADI, Guillory JE and Hancock JT (2014) 'Experimental evidence of massive-scale emotional contagion through social networks', *Proceedings of the National Academy of Sciences* 111(24): 8788–8790. The above discussion of this study draws from the work of the (U.S.) Council for Big Data, Ethics, and Society as summarised in Metcalf, Jacob, Emily F. Keller, and Danah Boyd 2017, 'Perspectives on Big Data, Ethics, and Society', Council for Big Data, Ethics, and Society http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/.

- sustaining **consumer trust:** 'spooky' to more than a few outlier customers will usually be bad business or unsustainable government policy;
- **fairness** - in the sense of social equity;
- **individual dignity** – treating each person as a private actor equally unless there is a compelling reason to 'discriminate' by treating a particular individual differently; or
- **moral philosophy** or **ethics**?[11]

Discussion of fairness and ethics within many businesses and governments is often as fuzzy as a philosophy tutorial of first year undergraduates. Recognising this, the debaters usually move on to address more concrete issues, such as *how would this be reported in the tabloid newspapers and on the 6:30pm television news?* and *will this increase profit margin (or reduce cost of delivery of government services?*. Let us see if we can find a way to make an ethics based discussion more concrete.

7 Discrimination

Discrimination by sellers as between buyers is probably older than homo sapiens, going back to first barter by humanoids (or by any other species intelligent enough to so discriminate).

The first vexed question is as to when discrimination becomes unfair or otherwise unacceptable, or illegal.

This requires us to venture into the treacherous waters of the European-led debate about **customer profiling**[12] and differential treatment of (possibly unidentified) individuals based upon inferences as to their characteristics or likely behaviour. In the public policy context, this debate often translates as the acceptable limits to application of behavioural economics to nudge citizens towards making choices that are socially beneficial (as perceived by the public policy maker).[13]

Discriminating sellers evolved from bartering humanoids, to the merchant in the bazaar, to the London East End barrow boys, to new car showroom sales staff and to media publishers and sell-side platforms selling online media 'canvas' to advertisers. Many of us are familiar with being ignored by sales staff because we look too young/old/scruffy, and then chased around the web by tracking code based online advertisements. As the online behavioural advertising eco-system has demonstrated, it is possible to design privacy compliant data sharing architectures that enable prior browsing behaviour to be used for differential treatment of individuals based upon inferences as to their likely interests and preferences. This capability has become so pervasive and readily apparent that it is now understood by most, if not all, Internet users. Targeted online advertising generally now is considered by many users as benign, by some users as welfare increasing (viz. they see ads for products and services that they are more likely to be interested in), but by some users as spooky or annoying.

Discrimination becomes possible whenever price is opaque: for example, targeted online price discrimination based upon perceived price inelasticity of demand. Why offer the same price to an online shopper in Wilcannia (where there is no comparable local supplier) to the price offered to an online shopper in Alexandria? Why offer the same price to an online shopper in Point Piper to the price offered in Bexley North? Possibilities for discrimination

[11] For an interesting discussion of these issues, see Luciano Floridi and Mariarosaria Taddeo, 14 November 2016, *What is data ethics?* Phil. Trans. R. Soc. A 2016 374 20160360; DOI: 10.1098/rsta.2016.0360 http://rsta.royalsocietypublishing.org/content/374/2083.

[12] Profiling in general parlance is the process of construction and application of user profiles generated by data analysis. It enables refined price-discrimination, targeted servicing, fraud detection and extensive social sorting. It has a much more specific meaning under the GDPR. Profiling' is defined under article 4 of the GDPR as " any form of automated processing of personal data consisting of using those data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements." Article 20 of the GDPR sets out three criteria which may trigger the provisions on automated processing of personal data, namely, a decision has to be made about an individual; which has a legal effect for that individual or significantly affects him or her; and this decision is based solely on automated processing. If those three criteria are met, "the data subject shall have the right not to be subject to a decision (…) based solely on automated processing, including profiling, which produces legal effects concerning him or her or significantly affects him or her" (article 20, GDPR). The right "not to be subject" to automated decisions is generally interpreted as the right to object to such processing. For article 20 of the GDPR to apply, the decision must be based solely on automated processing: hence profiling as a precursor, input or guide to a human decision is not profiling within article 20.

[13] For example, Richard H. Thaler and Cass R. Sunstein, *Nudge: Improving Decisions about Health, Wealth and Happiness*, Yale University Press, 2008, Penguin Books 2009.

increase when the decision by the seller as to whether to offer at a particular prospective buyer, and on what non-price terms, also become opaque. The potentiality for data analytics to be used in this way is a step-change that requires consideration of social ethics.[14] For example, targeted offering of premium regulated insurance products: why offer a premium regulated policy to an inferred higher risk customer? This potentiality for more refined discrimination is now working its way through many industry sectors, and most particularly the insurance industry.[15]

Of course, often possibilities for discrimination are not given effect. Amazon could charge more when an order comes in from Wilcannia, but it elects not to do so. Prices at Woolworths and Coles stores do not vary widely across Australia (notwithstanding substantial variation in logistics costs) because these retailers elect to maintain broad price parity across Australia, presumably as a decision of corporate social responsibility. Broad social equity is the result - and possibly also these retailers' rationale. Other suppliers might elect to do otherwise.

Some non-benign applications of algorithms would be illegal. Examples include algorithms that in application effect illegal discrimination as to race, ethnicity, religious beliefs, health or sexual orientation. Law in Australia and many OECD countries prohibits discrimination based on (some or all of) gender, race, genetics, religion, sexual orientation, political affiliation or age. However, big data processes can predict all of those characteristics without actually looking for fields such as gender, race or age. Genotypes, particularly those related to physical characteristics, are increasingly powerful predictors of many outcomes, yet relatively few countries today have genetic discrimination laws. And as Cathy O'Neill's recent book *Weapons of Maths Destruction*[16] powerfully illustrates, discriminatory effects may often flow from benign application of non-transparent algorithms – and algorithms will be increasingly non-transparent as machine learning progresses.[17]

8 Profiling in the European Union

In the European Union, the developing law on profiling tries to make a distinction between innocuous intelligence gathering and potentially dangerous exploitation of our digital identities. But the line that divides the two extremes is neither thin nor precise. The General Data Protection Regulation (**GDPR**) distinguishes between what could be called 'common profiling', which involves analysing or predicting aspects of someone's life, and a narrower type of profiling that produces legal effects concerning an individual or significantly affects an individual. The second, a sub-set, is seen as 'high risk profiling' and is subject to specific rules under the GDPR, including greater transparency, the right for affected individuals to challenge decisions and the obligation to undertake a data protection impact assessment.

It is easy to identify some cases of high risk profiling: the automatic refusal of an online credit application, or rejection of an online job application by an e-recruiting site, without human decision making. But what about other, less obvious examples? What might relevant effects or outcomes be? In the different but perhaps analogous context of working out when information can be considered to be personal information "about an individual" in circumstances where only indirect identifiers are present, the group of European privacy regulators known as the Article 29 Working Party in *Opinion 4/2007 on the concept of personal data* suggests that any of a content element, a 'purpose' element or a 'result' are relevant.

The 'content' element is present in those cases where information is given about a particular person who is identifiable by the recipient, regardless of any purpose on the side of the data source or intermediate discloser of data.

---

[14] See Karen Yeung, *Algorithmic Regulation: A Critical Interrogation*, Kings College London Dickson Poon School of Law, Legal Studies Research Paper Series: Paper No. 2017-27, forthcoming in Regulation & Governance available at http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1748-5991

[15] A useful brief introduction is provided by actuary Guy Thomas, 'How risky are you? Why insurance works better with less discrimination', The Conversation, 15 July 2017 https://theconversation.com/how-risky-are-you-why-insurance-works-better-with-less-discrimination-79558?. For more comprehensive reviews, see U.S. National Association of Insurance Commissioners (NAIC), *Casualty Actuarial and Statistical (C) Task Force Price Optimization White Paper*, 19 November 2015, available at http://www.naic.org/index.htm; also U.K. Financial Conduct Authority, *Feedback Statement on Call for Inputs: Big Data in retail general insurance,* FS/16 September 2016 https://www.fca.org.uk/publication/feedback/fs16-05.pdf. For a broad review, see U.K. Financial Conduct Authority, *Price discrimination and cross-subsidy in financial services*, Occasional Paper no. 22 September 2016, available at www.fca.org.au.

[16] Cathy O'Neill, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Random House, 2016; see also Cathy O'Neill, '*How can we stop algorithms telling lies?',* The Guardian, 16 July 2017 https://www.theguardian.com/technology/2017/jul/16/how-can-we-stop-algorithms-telling-lies?CMP=Share_iOSApp_Other.

[17] See further Jenna Burrell, 'How the machine 'thinks': Understanding opacity in machine learning algorithms', *Big Data & Society*, January 5, 2016, DOI: https://doi.org/10.1177/2053951715622512

A 'purpose' element can be considered to exist when the data is used or are likely to be used, taking into account all the circumstances surrounding the precise case, with the purpose to evaluate, treat in a certain way or influence the status or behaviour of an individual.

A 'result' element can be considered to exist where use is likely to have an impact on a certain person's rights and interests, taking into account all the circumstances surrounding the precise case. It should be noted that it is not necessary that the potential result be a major impact. It is sufficient if the individual may be treated differently from other persons as a result of the processing of such data.

These criteria are useful in working out whether to consider a possible outcome through a fairness or ethics frame.

9 Practical fairness/ethics/trust

What should an fairness/ethics/trust frame look like?

The UK Cabinet Office's *Data Science Ethical Framework*[18] suggests a 'six step approach":

(a) Start with clear user need and public benefit

(b) Use data and tools which have the minimum intrusion necessary

(c) Create robust data science models

(d) Be alert to public perceptions

(e) Be as open and accountable as possible

(f) Keep data secure.

As to **benefit**, the Data Science Ethical Framework states:

> Both the discovery and application phases require an organisation to define the benefits that will be created by the analytics and should identify the parties that gain tangible value from the effort. The act of big data analytics may create risks for some individuals and benefits for others or society as a whole. Those risks must be counter-balanced by the benefits created for individuals, organisations, political entities and society as a whole. Some might argue that the creation of new knowledge is a value-creating process itself. While big data does not always begin with a hypothesis, it usually begins with a sense of purpose about the type of problem to be solved. Data scientists, along with others in an organisation, should be able to define the usefulness or merit that comes from solving the problem so it might be evaluated appropriately. The risks should also be clearly defined so that they may be evaluated as well. If the benefits that will be created are limited, uncertain or if the parties that benefit are not the ones at risk from the processing, those circumstances should be taken into consideration, and appropriate mitigation for the risk should be developed before the analysis begins.

As to **respectful** uses:

> Respectful relates directly to the context in which the data originated and to the contractual or notice related restrictions on how the data might be applied. The United States Consumer Privacy Bill of Rights speaks to data being used within context; European law discusses processing not incompatible to its defined purpose; and Canadian law allows for implied consent for evolving uses of data. Big data analytics may affect many parties in many different ways. Those parties include individuals to whom the data pertains, organisations that originate the data, organisations that aggregate the data and those that might regulate the data. All of these parties have interests that must be taken into consideration and respected. For example, a specialised social network might display data pertaining to individuals that they would not expect to be used for, or would

[18] The UK Cabinet Office, *Data Science Ethical Framework* , ver. 1.0, May 2016, available at https://www.gov.uk/government/publications/data-science-ethical-framework. See also UK Information Commissioner's Office, 2017, *Big data, artificial intelligence, machine learning and data protection*, https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf.

be inappropriate for, employment related purposes. Organisations using big data analytics should understand and respect the interests of all the stakeholders involved in, or affected by, the application. Anything less would be disrespectful.

As to fairness:

The analysis of fairness needs to look not only at protecting against unseemly or risky actions but also at enhancing beneficial opportunities. Human rights speak to shared benefits of technology and broader opportunities related to employment, health and safety. Interfering with such opportunities is also a fairness issue.

In conducting this fairness assessment, organisations should take steps to balance individual interests with integrity.

A number of think tanks have endeavoured to make discussions as to business ethics more tractable.

In the United States, Marty Adams and Peter Cullen and their team at the Information Accountability Foundation (IAF) have been doing important work in developing an assessment framework for accountability and governance of big data projects. This work includes development of templates for project assessment and testing of those templates in different regulatory environments, including Canada.[19] The writer has found these templates practical and useful in broadening the scope of data analytics project assessment beyond (but also including) privacy impacts. In particular, the IAF articulates "Ethical Data Use Principles' and a methodology for applying them.[20] The Principles (in outline) are that data use should be Beneficial, Fair, Respectful and Just, Transparent and Autonomy Protecting and performed with appropriate Accountability with a Redress Provision. The approach is envisaged to be part of a 'Comprehensive Data Impact Assessment', effectively a much expanded PIA.

The writer considers that the IAF approach, although a big step forward from classic PIAs, requires further development.

Firstly, the IAF approach does not take account of the phasing of data analytics projects and may create deadweight costs in upfront evaluation of alternative possible outputs and outcomes that the discovery phase proves unlikely or unattainable.

Secondly, the IAF approach does not focus attention at the front end upon privacy protective management of data inputs.

Thirdly, the IAF approach assumes that the same reviewers are appropriate for both the privacy and the ethics components. The IAF templates provide a framework to ensure relevant considerations are canvassed in a systematic way, but do not really assist in answering the hard questions of what is unfair, inequitable or corrosive of consumer trust. Adoption of the IAF framework or other good practice frameworks will not guarantee the right answers, just as following processes and checklists to comply with quality control standards in production of a product will not guarantee that each product is of acceptable quality. In the writer's view, any framework must be operationalised through practical questions asked to a group empowered to represent a range of views as to fairness. It is not possible to objectively assess what is fair, but a properly selected and empowered reference group may, like a jury, collectively develop a useful 'wisdom of the crowds' proxy by the accumulation of various subjective views. To ensure that the room is not an echo chamber for a 'go-forward-bravely approach', the reference group should include designated consumer or citizen advocates. By the nature of the subject matter the ethical component requires a number of different viewpoints to be engaged through discussion, often informed by consideration of specific context and possible uses as only articulated later in a data analytics project. In any event,

[19] Information Accountability Foundation, 2014 *A Unified Ethical Frame for Big Data Analysis* http://informationaccountability.org/publications/a-unified-ethical-frame-for-big-data-analysis/; Information Accountability Foundation (Marty Adams and Peter Cullen), February 2017, *Canadian Assessment Framework*, Information Accountability Foundation, http://informationaccountability.org/publications/. See also Peter Cullen, *The Need for An Ethical Framework*, posted 24 January 2017 at http://informationaccountability.org/the-need-for-an-ethical-framework/, Martin Abrams, *Comprehensive Data Impact Assessments Set the Stage for Accountability 2.0* posted 2 March 2017 at http://informationaccountability.org/comprehensive-data-impact-assessments-set-the-stage-for-accountability-2-0/.

[20] Summarised in IAF, *'Decisioning Process, Risk-Benefits Analysis Tool for Data Intensive Initiatives - Achieving Legal, Fair and Just Use of Data & Appropriate Individual Engagement'*, November 2016, available at www.informationaccountability.org. See also the other IAF publications referenced at the end of this paper.

there is currently no academic or societal consensus about how to tackle these issues. Consideration of these issues must also be culturally aware. By way of some broad generalisations:

- there are important generational divides as to what is private and what is fair, and what is neither,
- the views of Muslim Australians and First Australians may differ significantly from those Australian citizens of 4th generation Anglo-Irish settler descent, and
- views of many Australian citizens differ significantly from American, German, Dutch and Chinese citizens.

Assessments of the reservoir of 'trust' from which business and government to draw are also shift significantly over time. An assessment by an Australian Federal, State or Territory government agency in July 2017 might well be different to an assessment pre the 2016 global decline in citizen trust in politicians (and possibly also, but much more contestably in Australia, government and government agencies). In the Australian context, the recent cavalcade of Federal Government information handling crises – the Census, then Medicare/PBS, then Blood Service, then Centrelink - has fairly been characterised as "*when it comes to trust, a good offence is your worst defence*".[21]

So how do we best facilitate a broadly based ethics/fairness/trust debate in assessment a data analytics project? A more multifaceted ethics approach has been developed by the Markkula Center for Applied Ethics at Santa Clara University[22]. The Markkula Center notes:

- There is no one ethical approach that works. Different approaches can be assimilated into a single assessment.
- Making good ethical decisions requires a trained sensitivity to ethical issues and a practiced method for exploring the ethical aspects of a decision and weighing the considerations that should impact our choice of a course of action. Having a method for ethical decision making is absolutely essential. When practiced regularly, the method becomes so familiar that we work through it automatically without consulting the specific steps.
- Ethics is not the same as feelings. Feelings provide important information for our ethical choices. Some people have highly developed habits that make them feel bad when they do something wrong, but many people feel good even though they are doing something wrong. And often our feelings will tell us it is uncomfortable to do the right thing if it is hard.
- The more novel and difficult the ethical choice we face, the more we need to rely on discussion and dialogue with others about the dilemma. Only by careful exploration of the problem, aided by the insights and different perspectives of others, can we make good ethical choices in such situations. So committees or councils may synthesise different ethical approaches to enable a balanced outcome.

The Markkula Center has developed *A Framework for Ethical Decision Making*, which can be summarized as follows:

(a) *Recognize an Ethical Issue*

Could this decision or situation be damaging to someone or to some group? Does this decision involve a choice between a good and bad alternative, or perhaps between two "goods" or between two "bads"?

Is this issue about more than what is legal or what is most efficient? If so, how?

(b) *Get the Facts*

[21] Mathew Beard, 'When it comes to trust, a good offence is your worst defence', *The Mandarin*, 6 March 2017, http://www.themandarin.com.au/76454-high-price-to-pay-to-correct-the-public-record/#.WL4wJylRXjQ.mailto

[22] See the materials available at *Readings in Big Data Ethics* at Markkula Center for Applied Ethics at Santa Clara University, https://www.scu.edu/ethics/internet-ethics-blog/internet-ethics-views-from-silicon-valley/readings-in-big-data-ethics.html.

What are the relevant facts of the case? What facts are not known? Can I learn more about the situation? Do I know enough to make a decision?

What individuals and groups have an important stake in the outcome? Are some concerns more important? Why?

What are the options for acting? Have all the relevant persons and groups been consulted? Have we identified creative options?

(c) *Evaluate Alternative Actions*

Evaluate the options by asking the following questions:

- Which option will produce the most good and do the least harm? (The Utilitarian Approach)
- Which option best respects the rights of all who have a stake? (The Rights Approach)
- Which option treats people equally or proportionately? (The Justice Approach)
- Which option best serves the community as a whole, not just some members? (The Common Good Approach)
- Which option leads us to act as the sort of person/entity we want to be? (The Virtue Approach)

(d) *Make a Decision and Test It*

Considering all these approaches, which option best addresses the situation?

- If we told people we respect, or told a television audience-which option we have chosen, what would they say? (The Sunshine Test)

(e) *Act and Reflect on the Outcome*

- How can my decision be implemented with the greatest care and attention to the concerns of all stakeholders?
- How did my decision turn out and what have I learned from this specific situation?

10 Bringing it all together

The writer suggest that that where fairness/ethics/trust concerns can be anticipated, the degree of effect upon individuals should be considered in a well-executed project outputs assessment and an outcomes assessment each conducted by an appropriately constituted reference group and applying the criteria suggested above. I deliberately use a different term to privacy impact assessment which might be conducted by an individual (appropriately experienced and impartial) privacy professional. PIAs have a narrower focus: for example, as to whether relevant data linkage is through processes establishing a properly controlled and safeguarded data eco-system, such that linked information may be considered secure and properly de-identified, albeit not fully anonymised, and all outputs risk assessed as low or remote re-identification risk. This does not address unfairness or otherwise unacceptable effects.

Project outputs or outcome assessments often will not be possible at the discovery phase of a project. Provided that data inputs and working data are properly privacy protected, adverse effects should only occur if outputs are generated from the application phase of a project. Often the various possible outputs and outcomes can be properly identified and scoped late in the conduct of the application phase of a data analytics project. Where there are reasonably foreseeable unfair or otherwise adverse effects of possible project outputs or outcomes upon individuals, the degree of effect upon individuals should be considered through a well-executed project outputs assessment conducted early in the application phase of any significant data analytics project. This project outputs assessment should be conducted wherever adverse outcomes upon some individuals are reasonably foreseeable,

regardless of perceived social benefits for many and other perceived social utility of a project, and regardless of whether personal information is being used in any part of the project.

A project outputs assessment should then be revisited by conduct of an outcomes assessment when the applications are developed and the outputs (be they insights, algorithms or processes or methodologies) are ready for application. The outcomes assessment often will be a straightforward review to confirm that the outputs assessment was sufficiently scoped to anticipate and address possible outcomes arising through application of the outputs. However, sometimes, the outcomes assessment process may lead to recommendations as to caveats or qualifications as to uses of outputs that should be notified to prospective users of those outputs, or as to publication to ensure appropriate transparency and accountability as to application of outputs (to address otherwise reasonably anticipated public concerns as to possible adverse applications and outcomes).

Requirements that engender and enable trust and effect fair outcomes should be sufficiently transparent as to be understood by stakeholders. These requirements should also be embodied in detailed contractual and legal obligations and associated work processes and procedures in order to ensure that the arrangements are reliable, as well as being appropriately privacy protective. These requirements are exacting, both to implement and to verify on an ongoing basis that the requirements are being met. Appropriately protective arrangements by design must ensure that processes and procedures anticipate and mitigate reasonably foreseeable risks of failures of processes through human error or oversight and other things that may go wrong.

The risks of failures of controls due to poor specification or monitoring are significant. The adverse consequences of failures may affect many projects beyond any particular project which suffers the failure: such is the interconnectedness of trust, and its loss, in an interconnected world. This is not an area for trial and error and iterative process improvements: it is important to get data linkage projects right from the start. That said, we now have a developing international consensus around good practice in risk management and risk responsive design in management of personal information and information security. We need to also ensure that where there are reasonably foreseeable unfair or otherwise adverse effects of a projects outcomes upon individuals, the degree of effect upon individuals is properly considered through well-executed, but not unduly cumbersome, project outputs and outcomes assessment.

Peter G Leonard

Principal, Data Synergies
LI https://www.linkedin.com/in/peleonard/
E pleonard@datasynergies.com.au

19 July 2017

Further reference

Standards

IEEE Project P7003, *Algorithmic Bias Considerations*, https://standards.ieee.org/develop/project/7003.html

ISO/IEC 38505-1:2017 *Information technology -- Governance of IT -- Governance of data -- Part 1: Application of ISO/IEC 38500 to the governance of data* https://www.iso.org/standard/56639.html

ISO/IEC 29100:2011 *Information technology — Security techniques — Privacy framework*, http://standards.iso.org/ittf/PubliclyAvailableStandards/c045123_ISO_IEC_29100_2011.zip

National Institute of Standards and Technology, NISTIR 8062 *An Introduction to Privacy Engineering and Risk Management in Federal Systems*, January 2017 https://doi.org/10.6028/NIST.IR.8062

Standards under development

IEEE P7000 *Model Process for Addressing Ethical Concerns During System Design* (Working Group in process)

IEEE P7001 *Transparency of Autonomous Systems* (Working Group in process)

IEEE P7002 *Data Privacy Process* (Working Group in process)

IEEE P7003 *Algorithmic Bias Considerations* (Working Group in process)

Frameworks

Australian Government National Health and Medical Research Council, *National Health National Statement on Ethical Conduct in Human Research (2007) - Updated May 2015* https://www.nhmrc.gov.au/guidelines-publications/e72

Australian Government National Health and Medical Research Council, *Human Research Ethics Committees (HRECs)*, https://www.nhmrc.gov.au/health-ethics/human-research-ethics-committees-hrecs

Australian Computer Society Data Taskforce (Dr Ian Oppermann), *Working Document Version 6.1 24th June 2017*

Clark, K. Duckham, M. Guillemin, M. Hunter, A. McVernon, J., O'Keefe, C. Pitkin, C. Prawer, S. Sinnott, R. Warr, D. Waycott, J. (2015) *Guidelines for the Ethical use of Digital Data in Human Research*, The University of Melbourne, Melbourne

British Academy and the Royal Society, *Data management and use: Governance in the 21st cent*ury, June 2017

Data Governance Australia, *draft Code of Practice* http://datagovernanceaus.com.au/dga-code-of-practice/

Centre for Epidemiology and Biostatistics at the Melbourne School of Population and Global Health, *Guidelines for the Ethical Use of Digital Data in Human Research*, March 2015

Council of Europe Directorate General of Human Rights and Rule of Law, *Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data*, Strasbourg, 23 January 2017

Alison Holt and Oxford Internet Institute, *Voluntary Code: Guidance for Sharing Data*,

IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems*, http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

IEEE, Prioritizing Human Well-being in the Age of Artificial Intelligence

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

Information Accountability Foundation (Marty Abrams and Peter Cullen), February 2017, *Canadian Assessment Framework*, Information Accountability Foundation http://informationaccountability.org/publications/

Information Accountability Foundation, 2014 *A Unified Ethical Frame for Big Data Analysis* http://informationaccountability.org/publications/a-unified-ethical-frame-for-big-data-analysis/

Markkula Center for Applied Ethics at Santa Clara University, May 2009, A *Framework for Ethical Decision Making*, https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/a-framework-for-ethical-decision-making/

Irina Raicu at Markkula Center for Applied Ethics at Santa Clara University, May 2015, *What is Internet Ethics?* https://www.scu.edu/ethics/focus-areas/internet-ethics/what-is-internet-ethics/

NSW Government Department of Finance, *Guidelines for sharing information between government agencies and the non-government sector* https://www.finance.nsw.gov.au/ict/guidelines-sharing-information-between-government-agencies-and-non-government-sector

Office of the Australian Information Commissioner, *Guide to big data and the Australian Privacy Principles: Consultation Draft* https://www.oaic.gov.au/engage-with-us/consultations/guide-to-big-data-and-the-australian-privacy-principles/

Office of the Australian Information Commissioner, *Guide to undertaking privacy impact assessments* https://www.oaic.gov.au/agencies-and-organisations/guides/guide-to-undertaking-privacy-impact-assessments

Office of the Australian Information Commissioner, *Health information and medical research* https://www.oaic.gov.au/privacy-law/privacy-act/health-and-medical-research

Office of the Australian Information Commissioner and Australian Government National Health and Medical Research Council, *Guidelines approved under Section 95A of the Privacy Act 1988* https://www.nhmrc.gov.au/guidelines-publications/e43

Office of the Australian Information Commissioner and Australian Government National Health and Medical Research Council, *Guidelines Under section 95 of the Privacy Act 1988* https://www.nhmrc.gov.au/guidelines-publications/e26

Office of the Australian Information Commissioner, *Use and disclosure of genetic information to a patient's genetic relatives under Section 95AA of the Privacy Act 1988 (Cth) – guidelines for health practitioners in the private sector* https://www.legislation.gov.au/Details/F2014L00244

Office of the Australian Information Commissioner, *Privacy management framework: enabling compliance and encouraging good practice* https://www.oaic.gov.au/agencies-and-organisations/guides/privacy-management-framework

Office of the Australian Information Commissioner, *Guidelines on Data Matching in Australian Government Administration* https://www.oaic.gov.au/agencies-and-organisations/advisory-guidelines/data-matching-guidelines-2014

UK Cabinet Office, *Data Science Ethical Framework*, ver. 1.0, May 2016, available at https://www.gov.uk/government/publications/data-science-ethical-framework

Useful books and articles

Peter Cullen, *The Need for An Ethical Framework*, posted 24 January 2017 at http://informationaccountability.org/the-need-for-an-ethical-framework/

Tanvi Desai, Felix Ritchie and Richard Welpton, 2016, *Five Safes: designing data access for research*, University of the West of England Bristol, Economics Working Paper Series 1601 http://eprints.uwe.ac.uk/28124/

Ariel Ezrachi and Maurice E. Stucke, , https://www.competitionpolicyinternational.com/the-rise-of-behavioural-discrimination/; also *Virtual Competition: the Promise and Perils of the Algorithm-Driven Economy*, Nov 2016

Federal Trade Commission, Remarks of Commissioner Terrell McSweeny, *Algorithms and Coordinated Effects*, University of Oxford Centre for Competition Law and Policy, Oxford, UK, May 22, 2017, https://www.ftc.gov/system/files/documents/public_statements/1220673/mcsweeny_-_oxford_cclp_remarks_-_algorithms_and_coordinated_effects_5-22-17.pdf

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

Financial Conduct Authority (U.K.), *Price discrimination and cross-subsidy in financial services*, Occasional Paper no. 22 September 2016 www.fca.org.au

Luciano Floridi and Mariarosaria Taddeo, 14 November 2016, *'What is data ethics?'*, Phil. Trans. R. Soc. A 2016 374 20160360; DOI: 10.1098/rsta.2016.0360 http://rsta.royalsocietypublishing.org/content/374/2083

Theo Forbath, Allison Schoop and Timothy Morey, '*Customer Data: Designing for Transparency and Trust*', Harvard Business Review May 2015 https://hbr.org/2015/05/customer-data-designing-for-transparency-and-trust

Michal S. Gal and Niva Elkin-Koren, '*Algorithmic Consumers*', Harvard Journal of Law & Technology, Volume 30, Number 2 Spring 2017

Gry Hasselbalch and Pernille Tranberg, *Data Ethics: The New Competitive Advantage*, 2017

Robert S. Kaplan and Anette Mikes, '*Managing Risks: A New Framework*', Harvard Business Review, 06/2016, https://hbr.org/2012/06/managing-risks-a-new-framework

Metcalf, Jacob, Emily F. Keller, and Danah Boyd, 2017, *'Perspectives on Big Data, Ethics, and Society'*, Council for Big Data, Ethics, and Society http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/

Metcalf, Jacob and Kate Crawford, '*Where are the human subjects in big data research? The emerging ethics divide*', Big Data and Society, Spring 2016

Cathy O'Neill, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Random House, 2016

Office of the Australian Information Commissioner, *Guide to undertaking privacy impact assessments* https://www.oaic.gov.au/agencies-and-organisations/guides/guide-to-undertaking-privacy-impact-assessments

UK Information Commissioner's Office, *Big data, artificial intelligence, machine learning and data protection*, https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

UK Information Commissioner's Office, *Overview of the GDPR*, section 8. Rights related to automated decision making and profiling; https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/rights-related-to-automated-decision-making-and-profiling/

UK Information Commissioner's Office, *Feedback request – profiling and automated decision-making*, https://ico.org.uk/media/about-the-ico/consultations/2013894/ico-feedback-request-profiling-and-automated-decision-making.pdf

US Executive Office of the President (White House), *Big Data: A Report on Algorithmic Systems, Opportunity and Civil Rights*, May 2016, https://www.whitehouse.gov/sites/default/files/.../2016_0504_data_discrimination.pdf

World Economic Forum, *Rethinking Personal Data: A New Lens for Strengthening Trust*, 2014, http://reports.weforum.org/rethinking-personal-data/