AHRC: Human Rights And Technology Issues Paper

July 2018

Submission:

Legal Implications Of Predictive Data

Claire Devenney and Christopher D Mills

Submission prepared as part of Juris Doctor studies at Flinders University

with

Tania Leiman,

JD Topic Coordinator, Flinders University

28 September 2018

Introduction / Overview

Big data refers to 'high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization.'[1] Rapidly increasing computing capacity now gives rise to the possibility of applying algorithms, machine learning and other forms of artificial intelligence to aggregate and analyse extremely large sets of both structured and unstructured data to generate new types of predictive data. This technological capacity has emerged only recently, creating new opportunities and new concerns. Strategic direction, policy and decisions in both public and private sectors increasingly adopt a data-driven, evidence-based approach. Predictions generated by big data analysis might thus be used to inform decisions at both community and individual levels about provision of access to resources such as insurance services, banking services, employment, accommodation, credit, educational opportunities, health care and government benefits. It may also form the basis for decisions in the criminal justice system[2], migration, and licensing. India's Aadhaar program, the world's largest biometric database containing the personal details of an estimated 1.2 billion Indian citizens, is just one example of the huge amount and type of data which may now be collected. [3]

Concerns arise where 'predictions' generate considerable specificity regarding existing or future risks to cohorts or to individuals, and those individuals are then discriminated against on the basis of that prediction.

Big data is being used to advance medical research. Data is collected from widely varying sources including administrative, clinical registers, electronic health records, biometric data,

[1] Michael Matzer, What Exactly Is Big Data?< https://news.sap.com/2012/11/what-exactly-is-big-data/> ; Bernard Marr, *What is Big Data? A Super Simple Explanation for Everyone* (March 2018) Bernard Marr & Co < https://www.bernardmarr.com/default.asp?contentID=766>.
[2] http://www.equivant.com/solutions/case-management-for-supervision
[3] https://edition.cnn.com/2018/09/26/asia/india-aadhaar-ruling-intl/index.html

'meteorological, climatic, and socio-political data'[4]. Data may also include individual's food and nutrition information obtained from health devices,[5] fitness devices, social media and internet search history[6].

The new predictive data is not personal information within the definition of the Privacy Act such as nor can it be linked information gained from a person. It is not detail about a retrospective health condition, nor is it test results of genetic biological material. It is not data from genetic testing, or mandatory screening or reporting of existing illness, injury or personal characteristics.

Big data analysis then generates new predictive data regarding health or medical risks, assisting in planning for appropriate resources allocation, disease prediction, treatment and population management.[7] Examples include predictions of outbreaks of denge fever in the Philippines[8] and,

---

[4] APL scientists use data to predict disease outbreaks weeks in advance < https://hub.jhu.edu/2013/03/21/apl-data-model-disease/>; Anna L. Buczak, Benjamin Baugher, Erhan Guven, Liane C. Ramac-Thomas, Yevgeniy Elbert, Steven M. Babin, andSheri H. Lewis, '

[5] Chloe Aiello, *Under Armour says data breach affected about 150 million MyFitnessPal accounts* (29 March 2018) CNBC < https://www.cnbc.com/2018/03/29/under-armour-stock-falls-after-company-admits-data-breach.html>.

[6] Mining consumers' web searches can reveal unreported side effects of drugs, researchers say, <https://med.stanford.edu/news/all-news/2013/03/mining-consumers-web-searches-can-reveal-unreported-side-effects-of-drugs-researchers-say.html>

[7] Louisa Jorm, *The future of 'big data' in health and medical research* (24 August 2017) University of NSW <https://researchaustralia.org/wp-content/uploads/2016/03/Future-of-Big-Data-in-HMR-Jorm.pdf>.

[8] APL scientists use data to predict disease outbreaks weeks in advance https://hub.jhu.edu/2013/03/21/apl-data-model-disease/; Anna L. Buczak, Benjamin Baugher, Erhan Guven, Liane C. Ramac-Thomas, Yevgeniy Elbert, Steven M. Babin, and Sheri H. Lewis, 'Fuzzy association rule mining and classification for the prediction of malaria in South Korea' (2015) 15 *BioMed Central Medical Informatics and Decision Making* 47 doi: 10.1186/s12911-015-0170-6

malaria in South Korea;[9] risks of suicide,[10] pancreatic cancer,[11] psychosis and mental illness[12], future development of Parkinson's disease.[13]

This new predictive data can identify persons with specified characteristics or behaviours as being at increased risk of either suffering from an existing undiagnosed health condition or of some other future health event. It can also be used to build models that predict possible future physical and emotional behaviours and states.[14]

This new predictive data is a new type of data. It is may not be based any actual characteristics or events of a particular person, yet may be used to define current or future outcomes or opportunities for that person or for a cohort in which that individual belongs. The predictive data may describe a particular person very specifically, even though it does not include any of their own personal or private information. It may the existence in a particular person now or in future of conditions that person is not even aware of. It could therefore be thought of in some sense as 'private data', even though it is not private to that individual. The individual becomes 'labelled' with that piece of predictive data on the basis of characteristics that are similar or can be predicted using the Big Data from a given cohort.

---

[9] Anna L. Buczak, Benjamin Baugher, Erhan Guven, Liane C. Ramac-Thomas, Yevgeniy Elbert, Steven M. Babin, and Sheri H. Lewis, 'Fuzzy association rule mining and classification for the prediction of malaria in South Korea' (2015) 15 *BioMed Central Medical Informatics and Decision Making* 47 doi: 10.1186/s12911-015-0170-6

[10] Internet search engines as both aid, danger for people at risk of suicide < https://www.sciencedaily.com/releases/2014/07/140728080442.htm>

[11] John Paparrizos, Ryen W. White, Eric Horvitz, 'Screening for Pancreatic Adenocarcinoma Using Signals From Web Search Logs: Feasibility Study and Results' Journal of Oncology Practice Volume 12 / Issue 8 / August 2016 DOI: 10.1200/JOP.2015.010504;
Michael Reilly, Can You Really Spot Cancer Through a Search Engine,< https://www.technologyreview.com/s/601656/can-you-really-spot-cancer-through-a-search-engine/>

[12] James Kirkup, 'Google wants to monitor your mental health. You should welcome it into your mind' https://www.telegraph.co.uk/news/health/11961415/Google-wants-to-monitor-your-mental-health.-You-should-welcome-it-into-your-mind.html

[13] *Shamli & Sathiyabhama. Parkinson's Brain Disease Prediction Using Big Data Analytics* International Journal of Information Technology and Computer Science(IJITCS), 8(6).

[14] Sam Volchenboum, *Social Networks May One Day Diagnose Disease – But At A Cost* (26 June 2017) Wired < https://www.wired.com/story/social-networks-may-one-day-diagnose-disease-but-at-a-cost/>.

A clear tension exists between an individual right to privacy and the benefit to society as a whole from scientific and technological advancement and innovation. This tension is profound in medicine and healthcare, where the public equally expects to benefit from advancement whilst privacy is maintained. Ground-breaking work in Australia has already used aggregated but de-identified data to document important risks that would otherwise have gone un-noticed or been imperfectly understood. These include risks from folic acid deficiency in pregnancy, risks of blood clots from long-haul air travel, and the increased cancer risks after CT scans (medical X-rays) in childhood.[15] The data forming the basis for this important evidence-based medical and health research places individuals contributing to those data sets at some risk of having their privacy breached or their identity discovered. Those risks have to date been managed by the Privacy Act 1988 (Cth) and state privacy frameworks. The capacity to analyse very large structured and unstructured data sets now gives rise to a new risk: that predictive data may be used adversely against that individuals or cohorts predicted to be at risk, including to discriminate against them.

Privacy and Human Rights

Article 12 of the *UN Declaration of Human Rights 1948* and Article 17 of the *International Covenant on Civil and Political Rights 1967* assert the right to be free from arbitrary or unlawful interference from 'privacy, family, home or correspondence' and from 'attacks upon … honour and reputation'. This is also enshrined in Section 13(a) of the *Charter of Human Rights and Responsibilities 2006* (Vic) s. 13(a) and section 12 of the *Human Rights Act 2004* (ACT).

[15] Fiona Stanley, Terry Nolan & John Mathews, *Census: the social compact on de-identified data protects privacy* (8th August 2016) The Mandarin < https://www.themandarin.com.au/68569-census-the-social-compact-on-deidentified-data/>.

The World Medical Association (WMA)'s *Declaration on Ethical Considerations Regarding Health Databases and Biobanks,*[16] recognises the public value of health research,[1] and the importance of privacy and confidentiality in allowing individuals to 'exercise control over their personal data and biological material'.

There is no common law right to privacy in Australia. The *Privacy Act 1988* (Cth) *Australian Privacy Principles* protect an individual's 'personal information', which is defined as *'Information or an opinion about an identified individual, or an individual who is reasonably identifiable:*

a. *whether the information or opinion is true or not; and*
b. *whether the information or opinion is recorded in a material form or not.'*[17]

Health information is defined as information or an opinion about the health or disability of an individual, ones expresses wishes about future provision of health services, a health service provided to an individual or other personal information collected in providing a health service or in connection with the donation (or intended) of one's body parts. [18] Genetic information on an individual was added to this definition was more recently added.[19]

Sensitive information is information or an opinion about an individuals racial, politics, religion, sexuality, criminal record, health information, genetic information, biometric information or philosophical beliefs, memberships.[20]

---

[16] World Medical Association, *Ethical Considerations Regarding Health Databases and Biobanks* (*Declaration of Taipei*), adopted by 53rd WMA General Assembly (October 2002) as revised by 67th WMA General Assembly (October 2016) ('*Declaration of Taipei*').

[17] Privacy Act 1998(Cth) s.6
[18] Ibid.
[19] Australian Law Reform Commission and Australian Health Ethics Committee, *Essentially Yours: The Protection of Human Genetic Information in Australia*, ALRC 96 (2003).
[20] Australian Law Reform Commission and Australian Health Ethics Committee, *Essentially Yours: The Protection of Human Genetic Information in Australia*, ALRC 96 (2003).

Each of these definitions relate specifically to 'information about an identified individual'. Whilst the definition in the Act of health information does consider data for the use of predictive health of an individual, the definition again relates to information about an individual.[21]

"Health information' is only protected under section the Privacy Act if it is also personal information or if it is '*genetic information about an individual in a form that is, or could be, predictive of the health of the individual or a genetic relative of the individual'*.[22]

While predictive data might be described as opinion about an individual who is reasonably identifiable, whether the opinion is true or not, if it only at-risk cohorts are reasonably identifiable, not individuals, then it will not engage the protections contained in the Australian Privacy Principles. The 'reasonably identifiable test' also raises interesting new challenges in the context of increased computing capacity, powerful data analytics tools and machine learning. What does reasonable mean in this context? Big data analysis of genetic information now make possible predictions about the future health of individual A even when genetic information about individual A is not part of the dataset, thus not 'about an individual', and so would also not engage the Australian Privacy Principles protections.

The Australian Privacy Principles also protect 'sensitive information', which includes 'health information about an individual', 'genetic information about an individual that is not otherwise health information', 'biometric information that is to be used for the purpose of automated biometric verification or biometric identification' or 'biometric templates'.[23] Even this definition may not cover predictive data about at-risk cohorts.

---

[21] Ibid.
[22] Privacy Act 1998(Cth) s.6FA
[23] Privacy Act 1988 (Cth) s.6

The *Privacy Act* only applies to identifiable data. Where data has undergone an appropriate and robust de-identification process it is not considered to be personal information and therefore the *Privacy Act* does not apply.[24]

The Australian Privacy Principles apply only to APP entities, and not to small business operators, registered political parties, State or Territory authorities or a prescribed instrumentality of a State.[25]

Most Australian states and territories also have some form of privacy protection framework in place:

Australian Capital Territory – *Privacy Act 1988* (Cth)
New South Wales – *Privacy and Personal Information Protection Act 1988*
Northern Territory – *Information Act 2002*
Queensland – *Information Privacy Act 2009*
South Australia – *Cabinet Administrative Instruction 1/89 2009*
Tasmania – *No laws*
Victoria – *Information Privacy Act 2000*
Western Australia – *No laws*

Data sharing legislation passed in SA (2015), NSW (2016), Victoria (2017) facilitates the sharing of data in the public sector and work is underway at the Commonwealth level to further enable data sharing amongst the government sectors.[26] The data sharing legislation referred to above has adopted the 'five safes, or trusted access principles, must be applied when sharing and/or using public sector data. In addition if data sharing is prohibited under various specified health provisions, an additional 'safe' has been prescribed that requires that the data not be shared or

[24] Office of the Australian Information Commissioner, 'De-identification and the Privacy Act' (2018) *Australian Government* 9.
[25] Privacy Act 1988 (Cth) s.6C
[26] Department of the Prime Minster and Cabinet, *New Australian Government Data Sharing and Release Legislation: Issues paper for consultation* (4 July 2018) Australian Government.

disclosed without the prior approval of the Minister for Health'.[27] The 'five safes' are safe projects, safe people, safe data, safe settings, and safe outputs. This has the potential to broaden the number and type of data sets that may be analysed to produce predictions about future health or behaviour. It may also give risk to increased risks of privacy breaches as data is further shared.[28]

Discrimination on the basis of Predictive Health Data

The following legislative framework exists across all Australian jurisdictions to protect against unlawful discrimination in various contexts and on various bases.

Federal Legislation	State Legislation
Age Discrimination Act 2004	Australian Capital Territory – *Discrimination Act 1991*
Australian Human Rights Commission Act 1986	New South Wales – *Anti-Discrimination Act 1977*
Disability Discrimination Act 1992	Northern Territory – *Anti-Discrimination Act 1996*
Racial Discrimination Act 1975	Queensland – *Anti-Discrimination Act 1991*
Sex Discrimination Act 1984	South Australia – *Equal Opportunity Act 1984*
	Tasmania – *Anti-Discrimination Act 1998*
	Victoria – *Equal Opportunity Act 2010*
	Western Australia – *Equal Opportunity Act 1984*.

Some discrimination is permitted by law in relation to health related information. For example, diagnosis of any patient with a disease listed in *National Health Security (National Notifiable*

[27] http://oda.sa.gov.au/content/five-safes
[28] Department of the Prime Minster and Cabinet, *New Australian Government Data Sharing and Release Legislation: Issues paper for consultation* (4 July 2018) Australian Government.

Disease List) Instrument 2018 (Cth) (‘*NNDS*’)[29] must be reported to the Department of Health under the Communicable Diseases Network Australia (CDNA) agreement, established under the National Health Security Agreement signed by all states and territories which enacts the *National Health Security Act 2007 (*Cth*)* (‘NHSA’)**.** For the purposes of Part 2 Public Health Surveillance, section 18 provides that ‘protected information’ in the NHSA is ‘personal information’ as defined in the *Privacy Act,* that has been collected in specified ways.*[30]* Part 2 sets out various permissible purposes for which information can be used, circumstances in which notifications about such information must be made and to whom, and various defences in relation to what would otherwise be breaches of confidentiality. The use of the word ‘personal information’ in the introduction to section 18 may mean that it will not cover predictive data.

Discrimination and predictions re risk of Infectious Diseases

It is not unlawful to discriminate against another person on the ground of the other person’s disability if: *(a) the person’s disability is an infectious disease; and (b) the discrimination is reasonably necessary to protect public health.*[31] In SA, any patient diagnosed with a Notifiable Disease in accordance with the NNDS may be ordered or directed to take remedial action at the behest of the Chief Public Health Officer. This includes certain controls such as being ordered to take an examination or test, counselling or if necessary, even be detained[32] but all such action must be undertaken subject to principles[33] which include being ‘*entitled to expect (a) to have his*

[29] https://www.legislation.gov.au/Details/F2018L00450
[30] *National Health Security Act 2007* (Cth) s 18.
[31] *National Health Security Act 2007* (Cth) s 18, *Disability Discrimination Act 1992* (Cth) s 48.
[32] *South Australian Public Health Act 2011* (SA) pt 10 div 2.
[33] Ibid s 14.

or her privacy respected and to have the benefit of patient confidentiality'.[34] Nevertheless, s 77 of the *South Australian Public Health Act 2011* (SA) ('*SAPHA*') provides that

> *(1) If—*
>
> *(a) the Chief Public Health Officer has reasonable grounds to believe that a person has, or has been exposed to, a controlled notifiable condition; and*
>
> *(b) the person is or has been the subject of 1 or more directions under section 75 and has contravened or failed to comply with a direction, or the Chief Public Health Officer considers that there is a material risk the person would not comply with 1 or more directions under that section if they were to be imposed; and*
>
> *(c) the Chief Public Health Officer considers—*
>
> *(i) that the person presents, or is likely to present, a risk to public health; and (ii) that action under this section is justified,*
>
> *then the Chief Public Health Officer may make an order under this section.*
>
> *…*
>
> *(4) An order under this section will be that the person submit to being detained at a specified place while the order is in forc*e.

Similar provisions also exist in all other states and territories in Australia.[35]

What if, however, a patient has not been diagnosed with a listed disease, but instead is *predicted* by big data analysis to either have the disease already, or be at significantly increased risk of contracting the disease? Predictions like these are already being made: for example a model that utilizes publicly available google search data to estimate flu like illness activity.[36]

This gives rise to the possibility that a person could potentially be subject to the public health protective provisions of the SAPHA if the Chief Public Health Officer became aware of information

---

[34] Ibid s 14(5)(a).
[35] See eg. *Public Health and Wellbeing Act 2008* (Vic).
[36] Shihao Yang, Mauricio Santillana, S.C Kou, *Accurate estimation of influenza epidemics using Google search data via ARGO* (9 November 2015) US National Library of Medicine <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4664296/>.

that, due to certain characteristics of that person, that they were at extreme risk of having a Notifiable Disease as a result of a 'predictive data' algorithm.

It is not at all clear that the privacy of such individuals would be similarly protected or that there would be any obligations regarding confidentiality in relation to the predictive data. In particular, instruments and legislation that protects discrimination on the basis of disability explicitly would not cover this scenario related to predictive information regarding an infectious disease.

This may mean that a person who is identified as being at risk of suffering from an infectious disease may face discrimination even though they do not have an existing disease; may not subsequently contract the disease; were not aware that they were at risk; may not be aware that an algorithm was analsying data for that purpose; may never have consented to being part of any data analysis; and may not be in a position to challenge the algorithm used in the analysis.

Discrimination and prediction of likelihood of future pregnancy

Social media has been used to identify women who are pregnant[37] even before they or their families may know.[38] Analysis of big data sets may provide an even more accurate prediction that a particular individual may be pregnant (even before a positive pregnancy test result had been obtained) or may be attempting to get pregnant (even before they have admitted this desire to themselves).

[37] Abeed Sarker, Pramod Chandrashekar, Arjun Magge, Haitao Cai, Ari Klein, and Graciela Gonzalez. 'Discovering Cohorts of Pregnant Women From Social Media for Safety Surveillance and Analysis' (2017) 19(10) *Journal of medical internet Research* e361 doi: 10.2196/jmir.8164; https://www.cnbc.com/2014/04/09/big-data-knows-youre-pregnant-and-thats-not-all.html; https://adage.com/article/digital/facebook-pregnant/237073/
[38] https://www.npr.org/templates/transcript/transcript.php?storyId=440305167

The *Sex Discrimination Act 1984* (Cth) ('*SDA*') s 7(1) provides that

> *(1) For the purposes of this Act, a person (the **discriminator**) discriminates against a woman (the **aggrieved woman**) on the ground of the aggrieved woman's pregnancy or potential pregnancy if, because of:*
>
> *(a) the aggrieved woman's pregnancy or potential pregnancy; or*
>
> *(b) a characteristic that appertains generally to women who are pregnant or potentially pregnant; or*
>
> *(c) a characteristic that is generally imputed to women who are pregnant or potentially pregnant;*
>
> *the discriminator treats the aggrieved woman less favourably than, in circumstances that are the same or are not materially different, the discriminator treats or would treat someone who is not pregnant or potentially pregnant.*

Section 4B of the *SDA* defines a potential pregnancy as

> *(a) the fact that the woman is or may be capable of bearing children; or*
> *(b) the fact that the woman has expressed a desire to become pregnant; or*
> *(c) the fact that the woman is likely, or is perceived as being likely, to become pregnant.*

Given this wide meaning of 'potential pregnancy', predictive data is likely to fall within s 4B of the *SDA*. However, it may be very difficult to know who has access to large data sets, and what purposes they are being used for. If employers use such predictive analytics to screen potential workers, it may be extremely difficult to identify that this has taken place, or to determine what the algorithm used and whether any predicted outcomes were reliable or accurate. In these circumstances, women may be discriminated against on the basis that they fall within a cohort that is very interested in becoming pregnant or may be in the very early stages of pregnancy. without any prospect of discovering that this less favourable treatment has taken place.

Similar concerns might arise in the context of mental or other illness.[39] The definition of disability in the *Disability Discrimination Act 1992* (Cth) 'includes a disability that … presently exists;

[39] https://bigthink.com/21st-century-spirituality/can-social-media-predict-depression-and-ptsd; https://draxe.com/social-media-mental-illness/; http://www.abc.net.au/news/2018-09-27/ai-better-than-doctors-at-predicting-mental-health-outcomes/10311388

previously existed but no longer exists; may exist in the future (including because of a genetic predisposition to that disability); or is imputed to a person.' [40]

Discrimination, insurance and genetic data

Parallels can be drawn from the way the insurance industry in Australia uses genetic data.[41] Genetic data is inherently predictive: In contrast to the United States, Canada and the Council of Europe's *Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine* (the Oviedo convention) (signatories include Belgium, Austria, Denmark, France, Germany, Lithuania, Norway, Portugal, and Sweden),[42] the Australian life insurance industry is left by the Australian Government to self-regulate through the Financial Services Council.[43] This has led to a concerning lack of regulation over the use of genetic information by Australian insurers, with insurers are allowed to use genetic test results to discriminate against applicants for life, permanent disability, and income protection insurance with minimal independent oversight, transparency or opportunity for intervention. Predictive data which can draw on a much broader range of information and datasets than genetic data poses even greater risks in insurance and other contexts. Insurers, mortgage and other loan providers, employers, accommodation providers, and education providers all may seek predictive data regarding risks presented by applicants and approve or deny benefits, services or opportunities accordingly. Challenging those decisions will be extremely difficult, especially where human decision makers may not even be

[40] Disability Discriination Act 1992(Cth) s.4
[41] K Barlow-Stewart and D Keays, 'Genetic Discrimination in Australia' (2001) 8 *Journal of Law and Medicine* 250, 254–256.' (2001) 8 Journal of Law and Medicine 250; S Taylor et al, 'Investigating Genetic Discrimination in Australia: A Large-Scale Survey of Clinical Genetics Clients' (2008) 74(1) *Clinical Genetics* 20.
[42] M Otlowski, S Taylor and Y Bombard, 'Genetic Discrimination: International Perspectives' (2012) 13 *Annual Review of Genomics and Human Genetics* 433.
[43] S Taylor et al, 'Investigating Genetic Discrimination in Australia: A Large-Scale Survey of Clinical Genetics Clients' (2008) 74(1) *Clinical Genetics* 20.

aware how an algorithm works. This is particularly problematic as algorithms may embed inherent or explicit biases, as various reports have explored in relation to US criminal justice predictive analysis tool COMPAS.[44]

Conclusion

Legislative protection against discrimination, while robust in some contexts, does not adequately address the use of predictive data, which when used has an even greater ability to discriminate[45] as discrimination legislation is often silent, or discrimination based on certain characteristics is already explicitly excluded.[46]

The examples of infectious diseases and pregnancy outlined above suggest Australia's current privacy and discrimination framework will need review to ensure it adequately protects human rights in the context of the emerging technology of predictive data. In particular, such review must protect against discrimination against persons who fall within the description of a cohort identified as being at risk by predictive analysis. Given the capacity of new technological tools to enable collection and analysis of big datasets in ways that were previously not possible, and the potential benefits and risks to society, governments must ensure that regulatory frameworks better protect against discrimination based on this new type of predictive data.

[44] http://www.equivant.com/solutions/case-management-for-supervision;
https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/;
http://advances.sciencemag.org/content/4/1/eaao5580;
https://epic.org/algorithmic-transparency/crim-justice/
[45] Anne Mignone, 'New Wine in Old Wineskins:The Implications of Australia's Privacy Laws for Biobanks' Honours LLBLP Thesis Flinders University, 2017.
[46] See eg. *Disability Discrimination Act 1992* (Cth) s 48