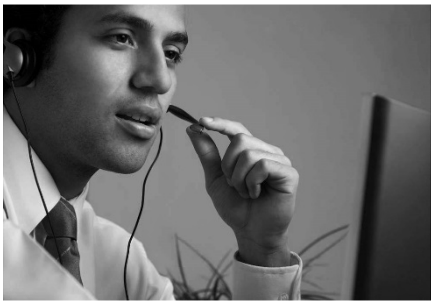
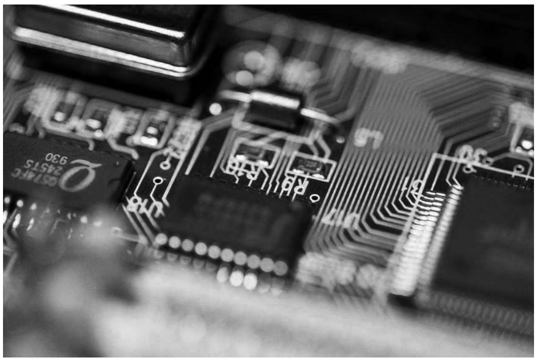
The Definition of Personal Information

Research Paper for the Office of the Australian Information Commissioner

17 February 2020





SalingerPrivacy

We know privacy inside out.

Executive Summary

Data is the lifeblood of the digital economy, and will increasingly power decision-making in all sectors of the economy.

Robust data protection regulation is necessary to achieve both consumer protection outcomes, and consistency of the playing field for industry. It will therefore be critical to ensure that the Privacy Act remains fit for its purpose of enabling effective regulation of personal information handling, in line with community and business expectations.

Through its Digital Platforms Inquiry, the ACCC found that the current definition of 'personal information' suffers from a lack of certainty around its coverage of technical data. The OAIC also raised the issue of whether inferred data is within scope, while a multiplicity of stakeholders expressed concerns that the Privacy Act – and in particular, the definition of 'personal information' - was not keeping up with the realities of the digital economy.

Whether or not any particular piece of data meets the definition of 'personal information' is a threshold legal issue for the operation of privacy law in Australia: the definition of 'personal information' determines the boundaries of what is regulated, and what is protected. Understanding the scope of what is meant by 'personal information' – and ensuring that that definition remains fit for purpose – is therefore a critical endeavour in privacy jurisprudence.

The challenges posed to the scope and reach of privacy laws come from many different directions: new technologies, new interpretations arising from case law, the increasing risks of re-identification, exponential growth in computing power, advances in fields like data analytics and cryptography, the phenomenon of data breaches, the influence of global debates, and new directions in statute law internationally.

Through the current definition of 'personal information', the Privacy Act regulates conduct only when a person is *identifiable*. However privacy harms can also arise from *individuation*: the ability to disambiguate or 'single out' a person in the crowd, even if that individual's 'identity' is not known. This poses a fundamental challenge for current privacy legal frameworks.

From the digital breadcrumbs we leave behind in the form of geolocation data shed from our mobile devices, to the patterns of behaviour we exhibit online as we browse, click, comment, shop, share and 'like', we can be tracked. Tracked, traced, monitored, surveilled; then profiled; and finally targeted … all without the party doing the tracking, profiling or targeting needing to know 'who' we are.

The digital environment has turned on its head the assumption that identifiability – in the sense of knowing a person's 'identity' - is only vector for privacy harm. Individuation must be anticipated by privacy laws as well.

These contemporary challenges lead us to conclude that the definition of 'personal information' in the Privacy Act no longer meets the needs of a privacy legal framework suitable for the digital age. The current definitions of 'personal information' and 'de-identified' in the Privacy Act fail to consider the privacy risks posed by individuation, or by attribute disclosure involving third parties.

Globally, other jurisdictions have more modern definitions, which clearly anticipate device identifiers, online identifiers and location data being used to identify – or at least 'single out' - individuals. Some privacy laws are broadening out the notion of 'identifiability' (or even abandoning it altogether) as the threshold element of their definition.

So as to enable clarity and consistency in the application of privacy law, and to protect against the potential privacy harms enabled by individuation, this Research Paper concludes that the Privacy Act should be amended, to incorporate a definition for the word 'identifiable':

> "(i) able to be identified, *or* (ii) able to be discerned or recognised as an individual distinct from others, regardless of whether their identity can be ascertained or verified"

This Research Paper further suggests that the test for identifiability should be that an individual will be considered "able to be discerned or recognised as an individual distinct from others":

> "if the individual, or a device linked to the individual, could (whether online or offline) be surveilled, tracked or monitored; or located, contacted or targeted; or profiled in order to be subjected to any action, decision or intervention including the provision or withholding of information, content, advertisements or offers; or linked to other data which is about or relates to the individual".

This additional layer to the test for identifiability aims to ensure that the scope of the regulation does not over-reach into technologies which do *not* pose risks of privacy harms, such as the use of sessional or load-balancing cookies which are necessary to make a website work, but which do not then continue to track the user.

By more explicitly embedding the concept of individuation within the core definitional element of 'identifiability', the Privacy Act can be modernised to reflect the reality of the digital economy, in line with other recent privacy laws such as the GDPR and CCPA.

This Research Paper's recommendations also include clarifying amendments to the definitions of 'personal information' and 'de-identified', as well as laying out a framework for how the privacy principles should work in practice with the revised definitions.

Anna Johnston
Principal | Salinger Privacy
17 February 2020

Contents

Why definitions matter

Whether or not any particular piece of data meets the definition of ‘personal information’ is a threshold legal issue for the operation of privacy law in Australia. Not only under the federal Privacy Act but also in State and Territory privacy laws, the definition of ‘personal information’ determines the boundaries of what is regulated, and what is protected, by the privacy principles which follow.

Privacy principles, tempered by exceptions for some scenarios, set out obligations on regulated entities for the handling of personal information, and create rights for individuals in relation to the personal information held about them. Data that is not ‘personal information’ is not subject to the same obligations, or the same protections – even if its collection or use is capable of doing harm to an individual.

Privacy laws around the world have the same effect. If data does not meet the threshold definition of ‘personal information’ (or its equivalent such as ‘personal data’ in Europe or ‘personally identifiable information’ in the USA), then the dataset can be released as open data, sold to other organisations, or used for a new purpose such as predictive analytics or to train a machine learning system, without legal limits or protections in relation to privacy.

Understanding the scope of what is meant by ‘personal information’ – and ensuring that that definition remains fit for purpose – is therefore a critical endeavour in privacy jurisprudence.

The current definition from s.6 of the Privacy Act is:

> ***"personal information"*** means 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.

Is the definition fit for purpose?

The last time the Privacy Act was reviewed, during the Australian Law Reform Commission's detailed inquiry over 2006-2008, the first iPhone had only just been released, and social media was in its infancy.

In the decade or so since, the world has changed dramatically. Estimates suggest that as at early 2020, 3.2 billion people have smartphones;[1] 2.5 billion people are active monthly users of Facebook;[2] and there are around 30 billion connected devices comprising the 'Internet of Things'.[3]

Peter Leonard describes the smartphone as "the clearest window to our souls and our demons":

> "As data collection and retention became cheaper over the decade and we moved to cloud services, we fed this new appendage more data, loaded and linked more applications, and placed it next to our pillows to monitor the quality of our sleep. … Today our mobile service provider usually knows less about what we are doing, and why we are doing it, than a myriad of mobile applications providers, the mobile operating system provider, and their contractors, many of whom we cannot even name."[4]

The question is whether our laws have kept up with these rapidly evolving technological advances, and their implications for our privacy – our autonomy, our self-determination and our solitude, our freedom of speech and freedom of association, and the freedom to live without discrimination or fear.

The challenges posed to the scope and reach of our privacy laws come from many different directions: new technologies, new interpretations arising from case law, the increasing risks of re-identification, exponential growth in computing power, advances in fields like data analytics and cryptography, the phenomenon of data breaches, the influence of global debates, and new directions in statute law internationally.

In 2019 the Australian Competition and Consumer Commission (ACCC) released the final report arising from its Digital Platforms Inquiry (DPI). The ACCC found that:

> "the Privacy Act needs reform in order to ensure consumers are adequately informed, empowered and protected, as to how their data is being used and collected. This will

[1] See https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

[2] See https://zephoria.com/top-15-valuable-facebook-statistics/

[3] See https://blogs.cisco.com/datacenter/internet-of-things-iot-data-continues-to-explode-exponentially-who-is-using-that-data-and-how

[4] Peter Leonard, "Beyond Data Privacy: Data "Ownership" and Regulation of Data-Driven Business", *SciTech Lawyer*, Vol 16(2), American Bar Association, 17 January 2020; available at https://www.americanbar.org/groups/science_technology/publications/scitech_lawyer/2020/winter/beyond-data-privacy-data-ownership-and-regulation-datadriven-business/

> increase trust in the digital economy and spur competition between businesses on the basis of privacy."[5]

In particular, the ACCC noted existing problems with legal uncertainty about the scope of the definition of 'personal information' under the Privacy Act in terms of its coverage of what the ACCC termed 'technical data'; and the need to offer Australian consumers more effective data protection standards to match those found in other jurisdictions, particularly Europe.

The ACCC concluded:

> "there are significant benefits in updating the definition of 'personal information' so that it covers the realities of how data is collected on individuals in the digital economy and to bring the Australian privacy regime into greater alignment with standards set by overseas data protection regulations."[6]

Recommendation 16(a) of the DPI Final Report was to update the definition of 'personal information' to clarify that it captures technical data such as IP addresses, device identifiers, location data, and any other online identifiers that may be used identify an individual.

Recommendation 17 related to a broader review of the Privacy Act, which the ACCC suggested should include consideration of whether inferred information should be protected by the Privacy Act.

Although the Law Council of Australia warned that as a legal threshold, any amendment to the definition of 'personal information' requires careful consideration and should involve minimal disruption, stakeholders making submissions to the ACCC on this point were broadly supportive of the need for reform.[7]

In December 2019 the Australian Government released its response to the DPI, including an implementation roadmap. The three responsible Ministers announced the Government's immediate commitment to a number of the DPI's recommendations, including to:

> "ensure privacy settings empower consumers, protect their data and best serve the Australian economy … through further strengthening of Privacy Act protections, subject to consultation and design of specific measures as well as conducting a review of the Privacy Act".[8]

[5] Australian Competition and Consumer Commission, *Digital Platforms Inquiry: Final Report*, June 2019, p.3; available at https://www.accc.gov.au/publications/digital-platforms-inquiry-final-report

[6] Australian Competition and Consumer Commission, *Digital Platforms Inquiry: Final Report*, June 2019, p.461; available at https://www.accc.gov.au/publications/digital-platforms-inquiry-final-report

[7] Law Council of Australia, Submission to the ACCC'S DPI, 15 February 2019, p.21; available at https://www.accc.gov.au/system/files/Law%20Council%20of%20Australia%20%28February%202019%29.PDF

[8] Australian Government, *Regulating in the digital age: Government Response and Implementation Roadmap for the Digital Platforms Inquiry*, December 2019, p.3; available at https://treasury.gov.au/sites/default/files/2019-12/Government-Response-p2019-41708.pdf

The Australian Government noted that:

> "While the benefits of digital services and technology are vast and will continue to grow, we must also be aware of, and respond appropriately to, the risks that are presented so that consumers and businesses have the confidence and capacity to engage in the digital world".[9]

The Government response and roadmap indicated the immediate commencement of consultation on Recommendation 16(a) in relation to the definition of personal information, as well as to shortly begin a broader review of the Privacy Act.

This Research Paper has been commissioned by the Office of the Australian Information Commissioner (OAIC) to inform the consultation on Recommendation 16(a).

The Productivity Commission, in its 2017 review of data sharing and availability, found:

> "The boundaries of personal information are constantly shifting in response to technological advances and new digital products, along with community expectations.
>
> The legal definition of personal information, contained in the *Privacy Act 1988* (Cth), has always had an element of uncertainty, and is managed by guidelines. In the face of rapid changes in sources and types of data, outcome-focused data definitions remain essential. But practical guidance (that data custodians and users can rely on) is required on what sorts of data are covered by the definitions."[10]

This Research Paper seeks to illuminate those questions, and pose suggested reforms by way of providing answers.

Challenges posed by the Internet of Things

Associate Professor Mark Andrejevic and Dr Mark Burdon have written about what they call the 'sensor society', in which the always-on interactive device is doubling as a tool for constant, passive data collection, even in our own homes.[11] Every connected device is capable of being a sensor, and monitoring its users. This turns privacy principles such as collection limitation, and limits on secondary use of data, on their head: "the function is the creep".[12]

[9] Australian Government, *Regulating in the digital age: Government Response and Implementation Roadmap for the Digital Platforms Inquiry*, December 2019, p.4; available at https://treasury.gov.au/sites/default/files/2019-12/Government-Response-p2019-41708.pdf

[10] Productivity Commission, *Data Availability and Use - Final Report*, May 2017, finding 3.4; available at https://www.pc.gov.au/inquiries/completed/data-access#report

[11] Mark Andrejevic and Mark Burdon, "Detection devices: how a 'sensor society' quietly takes over", *The Conversation*, 5 May 2014; available at http://theconversation.com/detection-devices-how-a-sensor-society-quietly-takes-over-26089

[12] Mark Andrejevic and Mark Burdon, "Defining the Sensor Society", *Television and New Media*, Vol 16(1), 2015, pp.19-36; available at https://espace.library.uq.edu.au/view/UQ:326402

The NSW Privacy Commissioner has reported to Parliament about the challenges that the Internet of Things poses for our privacy laws, as devices with unique identifiers push the boundaries of what is regulated as "personal information". Even devices which might be used by more than one person, such as smart meters within a home, are capable of identifying individuals within a group of users, simply from their patterns of behaviour.

> "We all now use a range of technological equipment each of which has its own identifying unit number and the ability to transmit electronic information without our express instigation. … This interconnection and our strong and close reliance upon our technological devices each of which is uniquely identifiable, will continue to raise challenges for the Act's definition of 'personal information'".[13]

Even the peak industry body IoT Alliance Australia (IoTAA) has argued that the definition of 'personal information' ought to be clarified or broadened so as to remove any doubt that the Australian Privacy Act regulates data collected by an "ever increasing range of sensing and actuating products". IoTAA noted that while each data collection might not be considered personal information in isolation, in combination the data can "yield highly personal information such as home occupancy and a wide range of behaviours".[14]

De-identification as a method of privacy protection is particularly difficult, if not impossible, in datasets featuring sensor data. Whether from a FitBit, an Amazon Echo, an Apple Watch or an internet-connected vehicle, the rich combination of location data and detailed behavioural data means one individual can be distinguished from millions of other individuals.[15]

Challenges posed by technical data

Data is powerful. In the analogue world, we already knew that the information on the outside of an envelope can be revealing and impactful on a person's privacy, even without seeing the contents inside. In the digital world, metadata is even more powerful. General Michael Hayden, former director of the NSA and the CIA, in a debate about telecommunications metadata, ascribed the ultimate value to 'technical data', or the type of data which some have argued is not 'private' or worthy of privacy protection: "We kill people based on metadata."[16]

[13] NSW Information & Privacy Commission, *Privacy Commissioner's Report under Section 61B of the Privacy and Personal Information Protection Act 1998*, February 2015, p.18; available at https://www.ipc.nsw.gov.au/privacy-commissioners-report-under-section-61b-privacy-and-personal-information-protection-act-1998

[14] IoT Alliance Australia, *Submission to the Australian Competition and Consumer Commission's Digital Platforms Inquiry*, 15 February 2019, p.1; available at https://www.accc.gov.au/system/files/Internet%20of%20Things%20Alliance%20Australia%20%28February%202019%29.PDF

[15] Scott Peppet, "Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent", *Texas Law Review*, Vol 93, 2014, p.129; available at https://texaslawreview.org/wp-content/uploads/2015/08/Peppet-93-1.pdf

[16] David Cole, "We Kill People Based on Metadata", *New York Review of Books*, May 10, 2014; available at https://www.nybooks.com/daily/2014/05/10/we-kill-people-based-metadata/

Technical data can be conceived of as the digital breadcrumbs we leave behind, whether in metadata about our communications, location data about our physical movements, or behavioural data about what we do online.

A number of privacy statutes explicitly refer to online identifiers and/or location data within their definitions of personal information or personal data:

- The General Data Protection Regulation (GDPR) covering the European Union (EU) includes online identifiers and location data as "identifiers", by reference to which an individual might be "identified, directly or indirectly", within its statutory definition of 'personal data' (Article 4). Recital 30 also notes that "Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags".

- The American *Children's Online Privacy Protection Rule* includes both geolocation information and any "persistent identifier that can be used to recognize a user over time and across different Web sites or online services. Such persistent identifier includes, but is not limited to, a customer number held in a cookie, an Internet Protocol (IP) address, a processor or device serial number, or unique device identifier" within its statutory definition of 'personal information'.[17]

- South African privacy law also explicitly includes "any identifying number, symbol, e-mail address, physical address, telephone number, location information, online identifier or other particular assignment to the person" within its statutory definition of 'personal information'.[18]

- Nigerian privacy law explicitly includes identifiers such as "MAC address, IP address, IMEI number, IMSI number, SIM" within its statutory definition of 'personal data'.[19]

However the scope of the Australian Privacy Act's definition of personal information is less clear, in terms of its coverage of technical data such as IP addresses, cookies, tracking pixels or metadata.

Dr Katherine Kemp and Dr Rob Nicholls submitted to the ACCC:

> "These types of information are regularly used by digital platforms to identify other information which is 'about an individual'. Expanding the definition of 'personal information' under the Privacy Act (to include technical data) is also in line with the definition of 'personal data' under the GDPR, which expressly includes online

[17] See https://www.ecfr.gov/cgi-bin/text-idx?SID=4939e77c77a1a1a08c1cbf905fc4b409&node=16%3A1.0.1.3.36&rgn=div5

[18] *Protection of Personal Information Act 2013*; available at https://www.justice.gov.za/inforeg/docs/InfoRegSA-POPIA-act2013-004.pdf

[19] *Nigeria Data Protection Regulation 2019*; available at https://nitda.gov.ng/wp-content/uploads/2019/01/NigeriaDataProtectionRegulation.pdf

> identifiers and location data. It is essential that Australia has a clear definition of ‘personal information’ which takes account of the realities of the digital age”.[20]

The Australian Privacy Foundation has critiqued the definition of personal information as “interpreted more narrowly than in Europe, (because) there is no indication of how to understand ‘indirect’ identification”. As a result, the Australian Privacy Foundation argues, information such as IP addresses and browsing history are treated as if they were not personal information subject to privacy protections.[21]

The position in Australia is actually less straight-forward than this critique would suggest.

A case involving metadata relating to mobile phone use turned on the meaning of the word ‘about’. At one stage the Administrative Appeals Tribunal (AAT) indicated that the complainant’s metadata was ‘about’ the device (his mobile phone), or about network connections, not about *him*, and therefore it did not constitute ‘personal information’.

Since the Privacy Commissioner lost the appeal against the AAT’s ruling, some have read the end result as suggesting that metadata was found not to be personal information. However the Federal Court was at pains to point out that it was *not* deciding one way or the other whether metadata actually met the definition of ‘personal information’; it was only deciding a point of law about the meaning of the word ‘about’. It did not then apply its reasoning to the facts in the particular case. (See further discussion below under *Case study: Metadata, identifiability and the meaning of ‘about’*.)

Given the Federal Court departed from the AAT ruling to clearly state that information and opinions *can* have multiple subject matters, the question of whether metadata does, or could, constitute ‘personal information’ remains entirely open.

The Note under the definition of ‘personal information’ in the Privacy Act states:

> Note: Section 187LA of the *Telecommunications (Interception and Access) Act 1979* **extends** the meaning of personal information to cover information kept under Part 5-1A of that Act. (Emphasis added.)

That section provides:

> 187LA Application of the *Privacy Act 1988*
>
> (1) The *Privacy Act 1988* applies in relation to a service provider, as if the service provider were an organisation within the meaning of that Act, to the extent that the activities of the service provider relate to retained data.

[20] Dr Katherine Kemp and Dr Rob Nicholls, from the Allens Hub for Technology Law and Innovation at the University of New South Wales, Submission to the ACCC on the DPI, 1 March 2019; available at https://www.accc.gov.au/system/files/Katharine%20Kemp%20%26%20Rob%20Nicholls%20%28March%202019%29.pdf

[21] Dr Monique Mann, “Privacy in Australia: Brief to UN Special Rapporteur on Right to Privacy”, Australian Privacy Foundation, 15 August 2018, p.5; available at https://privacy.org.au/wp-content/uploads/2018/08/Privacy-in-Australia-Brief.pdf

> (2) Information that is kept under this Part, or information that is in a document kept under this Part is taken, for the purposes of the *Privacy Act 1988*, to be personal information about an individual if the information relates to:
>
> (a) the individual; or
>
> (b) a communication to which the individual is a party.

There is an argument to be made that the effect of the drafting note in the Privacy Act is to suggest that were it *not* for s.187LA of the *Telecommunications (Interception and Access) Act 1979* (TIA Act), the types of metadata required to be kept by service providers would *not* be considered ‘personal information’.

The alternative view is that s.187LA of the TIA Act does in fact extend the scope of what would otherwise not be considered ‘personal information’, for the following reasons:

- ‘retained data’ under the TIA Act includes information about an “account, service or device”, the use of which could potentially only be attributable to an organisation or household (for example, a landline telephone service), and thus would likely not meet the definition of ‘personal information’ for that reason; and

- the language of the TIA Act is more inclusive than the Privacy Act, as it includes the broader concept of “if the information relates to” an individual (or, indeed, merely “relates to… a communication to which the individual is a party”, as opposed to the Privacy Act test which is only if the information is “about” an individual (a narrower subject-matter test).

The effect of s.187LA of the TIA Act is to encompass a protective notion about the privacy of communications, which recognises that information about devices can be a proxy for an individual (or, indeed, for households), who or which should be afforded some privacy rights and privacy protections, as a balance against the more intrusive aspects of the metadata retention scheme. It has the effect of bringing all ‘retained data’ under the umbrella of ‘personal information’, as a way of delivering those privacy rights and protections.

In our view, this does not necessarily suggest that metadata or technical data does not or cannot meet the definition of ‘personal information’ in its own right. Where telecommunications metadata, or indeed any other form of ‘technical data’, meets the current test in the Privacy Act – i.e. it is found to be ‘about’ an individual who is at least ‘reasonably identifiable’ – then it is ‘personal information’, regardless of the effect of s.187LA of the TIA Act.

Other Australian jurisdictions would appear to agree with this assessment. In a Queensland case, an IP address was found not to be personal information, because even in the course of a police investigation, neither Telstra nor the police service could determine its originating source, and thus they could not reasonably identify the individual concerned.[22] However the implication was that so long as an individual's identity could be reasonably ascertained (the test in the Queensland statute), then IP addresses would be covered. Guidance from the Office of the Victorian Information Commissioner (OVIC) states that "unique machine addresses for computers connected to the internet (for example, IP addresses), 'cookies' and other monitoring software" will be included, to the extent that they enable an individual's identity to be reasonably ascertained.

Challenges posed by location data

With the advent of mobile phones, telephony providers began to know where we were. With the shift to smartphones, that knowledge has spread well beyond just our phone providers; multiple smartphone apps use a mixture of GPS, Bluetooth and Wi-Fi signals to pinpoint locations whenever we carry our phones.

A global 'sweep' of more than 1,200 mobile apps by Privacy Commissioners around the world in 2014 found that three-quarters of all the apps examined requested one or more permissions; the most common was location.[23] Disturbingly, 31% of apps requested information not relevant to the app's stated functionality. A prominent example was the torch app which tracks users' precise location, and sells that data to advertisers.[24]

However it is not only apps we install on our mobile phones which can track our location. Bluetooth signals emitted by wearable devices can be collected by third parties; and venues such as shopping centres and airports (or, briefly, rubbish bins in London)[25] use the MAC addresses broadcast by devices to detect how populations are moving within a space, and to identify repeat visitors.[26]

Location data is highly granular. One study suggested that four points of geolocation data alone can potentially uniquely identify 95% of the population.[27] Mark Pesce, a futurist, inventor, educator and broadcaster, has described the geolocation data collected by and broadcast from our smartphones as "almost as unique as fingerprints".[28]

[22] *Lockyer Valley Regional Council and Queensland Police Service (311307)* at [26]; available at https://www.oic.qld.gov.au/decisions/lockyer-valley-regional-council-and-queensland-police-service.

[23] Office of the Privacy Commissioner of Canada, "From APP-laudable to dis-APP-ointing, global mobile app privacy sweep yields mixed results", 9 September 2014; available at https://www.priv.gc.ca/en/blog/20140909/

[24] See https://www.ftc.gov/news-events/press-releases/2013/12/android-flashlight-app-developer-settles-ftc-charges-it-deceived

[25] "U.K. bars trash cans from tracking people with Wi-Fi", CBS News, 12 August 2013; available at https://www.cbsnews.com/news/uk-bars-trash-cans-from-tracking-people-with-wi-fi/

[26] Jules Polonetsky and Elizabeth Renieris, Future of Privacy Forum Whitepaper: "*Privacy 2020: 10 Privacy Risks and 10 Privacy Enhancing Technologies to Watch in the Next Decade*", January 2020, p.4; available at https://fpf.org/wp-content/uploads/2020/01/FPF_Privacy2020_WhitePaper.pdf

[27] Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen & Vincent D. Blondel, 'Unique in the Crowd: The privacy bounds of human mobility', Scientific reports, March 2013, available at: https://www.nature.com/articles/srep01376?ial=1

[28] Mark Pesce was keynote speaker at the OAIC Business Breakfast for Privacy Awareness Week in May 2015; this quote is from the author's contemporaneous notes from the event.

Data showing where a person has been can reveal not only the obvious, like where they live and work or who they visit, but it may also reveal particularly sensitive information – such as if they have spent time at a church or a needle exchange, a strip club or an abortion clinic. Some app-makers claim they can even tell which floor of a building people are on.[29]

This Research Paper includes three case studies which demonstrate how location data has been used to identify (or at least, single out) individuals: see *Case study: Strava fitness data*, *Case study: NYC taxi trip data*, and *Case study: Myki public transport data*, below. In each case the dataset had purportedly been 'de-identified', but their release created the possibility of serious privacy harms including physical safety risks for some individuals in the dataset.

The Future of Privacy Forum has nominated location services and proximity tracking as one of their 'top 10' privacy risks for the 2020s. In particular, they note that because 5G signals have a shorter range than existing technology, they require more numerous and smaller cellular towers. This in turn will make the geolocation data generated about users even more granular than is currently available.[30]

Challenges posed by online tracking

Led by – but by no means exclusive to – the online behavioural advertising industry, entire ecosystems have been built around the use of online identifiers, which are deployed to track the behaviour of consumers and citizens across different sites, and over time.

Online identifiers can include:

- device identifiers (e.g. an IMEI mobile phone identification number, or a MAC address for devices which connect to networks or other devices using Wi-Fi or Bluetooth),
- device fingerprinting (e.g. information about the IP address, operating system, language settings and browser being used on a device)
- files embedded on devices (e.g. cookies, web beacons and tracking pixels), and
- digital platform user IDs used across different sites (e.g. using an existing Facebook or Google account to 'log in' to a third party site).

By linking a device to behaviour such as searches, queries, posts, browsing sites and purchases, the party doing the tracking can start to profile individuals, drawing inferences about their interests and preferences, behaviour and budget, and divide them into segments accordingly. The individual presumed to be the user of the device can then be targeted to receive a particular ad, or subjected to a decision such as differential pricing.

[29] David Pierce, "Location Is Your Most Critical Data, and Everyone's Watching", *Wired*, 27 April 2015; available at https://www.wired.com/2015/04/location/

[30] Jules Polonetsky and Elizabeth Renieris, Future of Privacy Forum Whitepaper: "*Privacy 2020: 10 Privacy Risks and 10 Privacy Enhancing Technologies to Watch in the Next Decade*", January 2020, p.4; available at https://fpf.org/wp-content/uploads/2020/01/FPF_Privacy2020_WhitePaper.pdf

Advertisers and data brokers will claim that none of this activity involves ‘personal information’ under the Australian definition, because they cannot ‘reasonably’ identify the individual. However the privacy impacts on the consumer are real:

> “segmentation of nonidentifiable individuals may lead to significant, systematic, and adverse effects on how some segments of individuals are treated relative to other segments. If I am a provider of airline reservations and I see a particular device repeatedly coming back for quotes, I may elect to increase the quoted price, inferring that no other provider is matching my quote. If I am an insurer and infer that a device is being used in a community where risks are typically higher, I may elect to quote a higher price than that which I offer to users known or inferred to be in lower risk locations.”[31]

Challenges posed by data analytics

Further, advances in data analytics, and the predictive capabilities of machine learning and artificial intelligence technologies, are also creating new challenges for the law’s ability to draw a bright line between what is ‘personal information’ and what is not.

The Office of the Victorian Information Commissioner has noted that:

> “the distinction between what is and is not considered to be ‘personal’ is being challenged by the increasing ability to link and match data to individuals, even where previously thought to be 'de-identified' or non-identifying to begin with.
>
> … a combination of seemingly non-personal information can become personal information when analysed or correlated. As the amount of available data increases, and technologies for processing and combining it improve, it becomes increasingly difficult to assess whether a given piece of data is ‘identifiable’; considering a piece of data in isolation is not compatible with AI technology, and is no longer a true reflection of whether it can be deemed ‘personal information’”.[32]

[31] Peter Leonard, “Beyond Data Privacy: Data “Ownership” and Regulation of Data-Driven Business”, *SciTech Lawyer*, Vol 16(2), American Bar Association, 17 January 2020; available at https://www.americanbar.org/groups/science_technology/publications/scitech_lawyer/2020/winter/beyond-data-privacy-data-ownership-and-regulation-datadriven-business/

[32] Office of the Victorian Information Commissioner, “Artificial intelligence and privacy: Issues Paper”, June 2018, p.9; available at https://ovic.vic.gov.au/resource/artificial-intelligence-and-privacy/

Challenges posed by re-identification

Time and time again, supposedly ‘de-identified’ datasets are re-identified.

For example, a 2000 study obtained publicly available health insurance information on Massachusetts state workers that was stripped of names, addresses, social security numbers and other ‘identifying’ information. Then PhD student Latanya Sweeney purchased state voter rolls for the city of Cambridge, including the name, ZIP code, address, sex and birth date of every registrant. The insurance data showed that there were six people in Cambridge born on the same day as the governor; half were men. The voter data allowed Sweeney to pinpoint the state's governor as the only one of those residing in a particular ZIP code in Cambridge. The corresponding health-insurance data included the governor's medical diagnoses and prescriptions.[33]

In 2006, search engine provider AOL released ‘anonymous’ web search records for 658,000 users. *New York Times* journalists linked search terms to identify users and contact them:

> “Trawling though the hundreds of searches made by Subscriber 4417749 for local estate agents and gardeners, through to ‘numb fingers’, ‘dog that urinates on everything’ and ‘60 single men’, they tracked down Thelma Arnold, a 62-year-old widow and pet-owner from Lilburn, Georgia. ‘My goodness, it's my whole personal life,’ she said as the reporter read AOL's search records to her. ‘I had no idea somebody was looking over my shoulder’.” [34]

Other examples have included a genetic database,[35] London’s bike-share scheme,[36] telephone metadata,[37] social network connections,[38] online ratings of television programs,[39] mobility data,[40] records of credit card transactions,[41] and taxi trip data.[42] In Australia, the examples have included MBS/PBS data,[43] and Myki public transport data.[44]

[33] Latanya Sweeney, “Computational Disclosure Control: A Primer on Data Privacy Protection”, January 2001; available at http://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/sweeney-thesis-draft.pdf
[34] Michael Barbaro and Tom Zeller Jr, “A Face Is Exposed for AOL Searcher No. 4417749”, *New York Times*, 9 August 2006; available at https://www.nytimes.com/2006/08/09/technology/09aol.html; Charles Duhigg, “How Companies Learn Your Secrets” *New York Times*, 16 February 2012; available at https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html
[35] Matt Fearer, “Scientists expose new vulnerabilities in the security of personal genetic information”, Whitehead Institute, 17 January 2013; available at http://wi.mit.edu/news/archive/2013/scientists-expose-new-vulnerabilities-security-personal-genetic-information
[36] James Siddle, “I Know Where You Were Last Summer: London's public bike data is telling everyone where you've been”, personal blog, 10 April 2014; available at https://vartree.blogspot.com/2014/04/i-know-where-you-were-last-summer.html
[37] Mudhakar Srivatsa & Mike Hicks, “Deanonymizing Mobility Traces: Using Social Networks as a Side-Channel”, *Computer and Communications Security*, October 2012: available at http://www.cs.umd.edu/~mwh/papers/GraphInfoFlow.CCS2012.pdf
[38] Arvind Narayanan, Elaine Shi & Benjamin Rubinstein, “Link Prediction by De-anonymization: How We Won the Kaggle Social Network Challenge”, 2011 International Joint Conference on Neural Networks, October 2011; available at: https://arxiv.org/pdf/1102.4374.pdf
[39] Arvind Narayanan & Vitaly Shmatikov, “Robust De-anonymization of Large Sparse Datasets”, *Security and Privacy*, May 2008; available at: https://www.cs.cornell.edu/~shmat/shmat_oak08netflix.pdf
[40] Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen & Vincent D. Blondel, “Unique in the Crowd: The privacy bounds of human mobility”, *Nature*, March 2013; available at https://www.nature.com/articles/srep01376?ial=1
[41] Yves-Alexandre de Montjoye, Laura Radaelli, Vivek Kumar Singh & Alex Pentland, “Unique in the shopping mall: On the reidentifiability of credit card metadata”, *Science*, Vol 347(6221), 2015; available at http://science.sciencemag.org/content/347/6221/536.full
[42] See *Case study: NYC taxi trip data*, below.
[43] Office of the Australian Information Commissioner, *Publication of MBS/PBS data: Commissioner initiated investigation report*, 23 March 2018, p.4; available at https://www.oaic.gov.au/assets/privacy/privacy-decisions/investigation-reports/publication-of-mbs-pbs-data.pdf
[44] OVIC, “Disclosure of myki travel information: Investigation under section 8C(2)(e) of the Privacy and Data Protection Act 2014 (Vic)”, 15 August 2019; available at https://ovic.vic.gov.au/wp-content/uploads/2019/08/Report-of-investigation_disclosure-of-myki-travel-information.pdf

In addition to the *deliberate* public release of datasets such as the above examples, student researchers from Harvard University have recently discovered that data breaches further increase the likelihood that even 'de-identified' datasets will lead to the re-identification of individuals.[45]

Conclusion

The definition of 'personal information' in the Privacy Act no longer meets the needs of a privacy legal framework suitable for the challenges of the digital age.

Other Australian federal laws are already supplementing the Privacy Act regime, with slightly broader terms to draw more information within their protective scope than would otherwise be possible if relying on just the definition of 'personal information' from the Privacy Act:

- the TIA Act protects data that could be about households rather than individuals, and adopts the broader 'relates to' test rather than the subject-matter 'about' test

- the new Consumer Data Right (CDR) scheme creates rights for a wider class of persons (both individuals and organisations), and also deliberately adopted the broader 'relates to' test rather than the narrow subject-matter 'about' test[46]

Globally, other jurisdictions have more modern definitions of 'personal information' (or its equivalent), which clearly anticipate device identifiers, online identifiers and location data being used to 'identify' individuals. Some privacy laws are broadening out the notion of 'identifiability' (or even abandoning it altogether) as the threshold element of their definition.

A global comparative analysis of each element of the definition of personal information is presented below. However, before hiking too deep into the trees of legal definitions, it is important to step back and see the forest.

The reason we have privacy laws is not to protect data; it is to protect people. It is the people who can be found in data, singled out because of data, tracked and even manipulated via data, who matter.

Privacy laws exist to protect people from privacy harms. It is because of the scope to do harm to people that some practices are deserving of regulation.

[45] Karl Bode, "Researchers Find 'Anonymized' Data Is Even Less Anonymous Than We Thought", *Vice*, 4 February 2020; available at https://www.vice.com/en_us/article/dygy8k/researchers-find-anonymized-data-is-even-less-anonymous-than-we-thought

[46] OAIC, *Complete version of the Draft Privacy Safeguard Guidelines (combined) Consultation draft*, October 2019, Chpt B, p.8; available at https://www.oaic.gov.au/assets/engage-with-us/consultations/draft-cdr-privacy-safeguard-guidelines/Complete-version-of-the-Draft-Privacy-Safeguard-Guidelines-combined.pdf

Privacy harms exist across a spectrum, and include:

- tangible or 'material' harms at one end (such as physical harm or threats of violence, stalking and harassment, identity theft, financial loss and psychological damage),
- intangible or 'moral' harms in the middle (such as reputational damage, "creepy inferences", humiliation, embarrassment or anxiety, loss of autonomy, discrimination and social exclusion), and
- abstract or 'social' harms at the other end (such as the threats to democracy, chilling effect on free speech, loss of trust and social cohesion posed by a 'surveillance society').[47]

Even though privacy obligations generally only arise in contexts or situations when a person is *identifiable*, privacy harms can arise from *individuation* (being disambiguated from the crowd), regardless of whether or not the person is also identifiable in a concrete or legally verifiable sense. In other words, a perpetrator can hurt someone without ever knowing who they are. This poses a fundamental challenge for current privacy legal frameworks.

Next steps

The Privacy Act needs a more suitable definition, fit for the digital age.

The remainder of this Research Paper examines, and makes recommendations about:

- the core elements of the definition of personal information
- the extent to which inferred information is (or should be) explicitly or implicitly included in the definition or explanatory memoranda
- the extent to which technical data is (or should be) explicitly or implicitly included in the definition or explanatory memoranda
- whether, in order to remain fit for purpose in the digital age, the definition should encompass the concept of individuation
- whether, in order to remain fit for purpose in the digital age, the definition should be extended to identifiable households and/or consumer devices
- how the Australian Privacy Principles (APPs) would work in practice with any expanded scope of the definition; and

[47] This spectrum of privacy harms is drawn from work by the former UK Information Commissioner, as well as the Future of Privacy Forum's paper, "Benefit-Risk Analysis for Big Data Projects", September 2014, available at www.futureofprivacy.org

- how any expanded scope would interplay with other proposed amendments to the Privacy Act such as increasing coverage of small businesses or employee records, or adding new data subject rights.

Any proposed reform must be mindful of the need for global consistency, which is beneficial for consumers, regulated entities and regulators alike; but must also ensure that the definition is ‘fit for purpose’ for Australian conditions now and into the future.

Elements of the definition

The definition of personal information has several elements:

- information or opinion
- about
- an individual
- who is identified or reasonably identifiable

The following discussion examines each of these elements in turn, and reviews how they are treated by the Privacy Act and privacy laws from comparable jurisdictions.

‘information or opinion’

The Australian definition of personal information includes “information or an opinion” about a person, “whether the information or opinion is true or not”.

By explicitly including opinion as well as information, the OAIC suggests that inferred data is already included in the definition. The OAIC’s guidelines provide the following examples of what will constitute ‘personal information’:

- “An opinion about an individual’s attributes that is based on other information about them, such as an opinion formed about an individual’s gender and ethnicity, based on information such as their name or their appearance. This will be personal information about the individual even if it is not correct.
- Information or opinion inferred about an individual from their activities, such as their tastes and preferences from online purchases they have made using a credit card, or from their web browsing history”.[48]

Guidance from the UK Information Commissioner’s Office (ICO) suggests that ‘special category’ data (such as sexuality or religion) which has been inferred from other data, rather than provided directly by the subject individual, will still be personal data, even if the inference is not made with certainty.[49]

[48] Office of the Australian Information Commissioner, *What is personal information?*, May 2017, p.6; available at https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information

[49] See https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category-data/

The Supreme Court of British Columbia has recognised the capacity for personal information to be inferred from metadata because of its ability, when combined, to track a person's activities and reveal their preferences.[50]

Two recently drafted privacy laws are explicit in their inclusion of inferred data:

- India includes "any inference drawn from such data for the purpose of profiling"; and
- the California Consumer Privacy Act (CCPA) explicitly includes, within its definition of personal information:
    - "purchasing or consuming histories or tendencies", and
    - "inferences drawn from" any of the other types of information enumerated in the definition, "to create a profile about a consumer reflecting the consumer's preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes".[51]

The CCPA defines "infer" or "inference" as "the derivation of information, data, assumptions, or conclusions from facts, evidence, or another source of information or data".[52]

This Research Paper suggests that for absolute clarity about the matters identified by the ACCC and others, the definition of personal information in the Privacy Act should include the phrase "whether the information or opinion is provided, collected, created, generated or inferred".

'about'

The OAIC has advised that:

> "Information is 'about' an individual where there is a connection between the information and the individual. This is ultimately a question of fact, and will depend on the context and the circumstances of each particular case."[53]

[50] *Abougoush v. Sauve*, 2011 BCSC 885 at [11]; available at https://www.bccourts.ca/jdb-txt/SC/11/08/2011BCSC0885.htm

[51] CCPA section 1798.140(o)(1), parts (D) and (K); the full text of the CCPA, as at 1 January 2020, is available at http://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=CIV&division=3.&title=1.81.5.&part=4.&chapter=&article=

[52] CCPA section 1798.140(m); the full text of the CCPA, as at 1 January 2020, is available at http://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=CIV&division=3.&title=1.81.5.&part=4.&chapter=&article=

[53] Office of the Australian Information Commissioner, *What is personal information?*, May 2017, p.6; available at https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information

> "Information will also be 'about' someone where it reveals or conveys something about them — even where the person may not, at first, appear to be a subject matter of the information."[54]

The tortured history of some case law suggests that interpretation of the word 'about' is not so straight-forward. (See a more detailed discussion below under *Case study: Metadata, identifiability and the meaning of 'about'*.)

A case involving metadata turned on the meaning of the word 'about', and the AAT indicated that the complainant's metadata was 'about' the device (his mobile phone), or about network connections, not about *him*. However on appeal the Federal Court found that information and opinions can have multiple subject matters.

There will however need to be some degree of connection that is not too remote. The Federal Court stated:

> "Information can have different degrees of connection with an individual and still be personal information. However, at some point, the connection between the information and a person will be too remote for the information to be personal information."[55]

NSW case law suggests that information about land, in terms of *land use*, works on a building on that land, or a building's interior design, can also be 'about' the individuals owning the land, or living or working there.[56]

The NSW Civil and Administrative Tribunal (NCAT) has stated for example that:

> "the fact that the information (at issue) relates to the land and the land use … does not prevent it being personal information within the wide and broad scope of the definition ... It records that a residence and nursery were observed as activities on the property owned by OS and his wife. While this undoubtedly refers to the land use, it also provides information about how OS uses his property and the activities pursued there".[57]

Similarly the Queensland Office of the Information Commissioner (OIC) has provided guidance that "information that the rates for a particular property have not been paid for a year is about the land, but it also reveals a fact about the owner, that they have not been paying the rates".[58]

[54] Office of the Australian Information Commissioner, *What is personal information?*, May 2017, p.7; available at https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information

[55] See *Privacy Commissioner v Telstra Corporation Limited* [2017] FCAFC 4 (19 January 2017) at [64] per Kenny and Edelman JJ., and see *Telstra Corporation Limited and Privacy Commissioner* [2015] AATA 991 (18 December 2015) at [43].

[56] See *WL v Randwick City Council (No 2)* [2010] NSWADT 84, *OS v Mudgee Shire Council* [2009] NSWADT 315 and *APV and APW and Department of Finance and Services* [2014] NSWCATAD 10.

[57] *OS v Mudgee Shire Council* [2009] NSWADT 315 at [23]; available at https://www.caselaw.nsw.gov.au/decision/549f6a3f3004262463a4f407

[58] See https://www.oic.qld.gov.au/guidelines/for-government/guidelines-privacy-principles/key-privacy-concepts/key-privacy-concepts-personal-information

Each of those cases involved information about what humans were doing to or on the land (i.e. how the land was being used by the individuals), as opposed to information about what external environmental factors might affect the land. Data about environmental conditions affecting a piece of land (e.g. that a particular property is prone to flooding or landslip) could be said to be of particular interest to certain individuals, such as the owner of that land or potential purchasers of the land, but it could not be said to be information 'about' any particular individual.

Other jurisdictions which also use the word 'about' include NSW, Victoria, Queensland and Tasmania; and New Zealand, Canada and Singapore.

The Federal Court of Appeal in Canada has elaborated that 'about' means that the information is not just the subject of something but also relates to or concerns the subject; although in that case the Court was also interpreting the French text of the statute which uses the word '*concernant*'.[59]

Other jurisdictions use different terms, such as:

- "relating to" (GDPR, South Australia, South Africa, Nigeria, Japan, Hong Kong; the Council of Europe's Convention 108 and 108+; and the 1980 *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* including the 2013 update)
- "about or relating to" (India)
- "regarding" (Brazil), or
- "to whom such information relates" (ISO 27701).

Some jurisdictions use no qualifier about the subject matter of the information at all. For example the Northern Territory definition is simply "Government information from which a person's identity is apparent or…".

The CCPA also eschews a separate subject matter test, in favour of a much broader scope:

> "information that identifies, relates to, describes, is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household".[60]

An argument against broadening the definition of 'about' is the need to maintain a balance between the rights of individuals and the legitimate interests of organisations and government agencies. However this balancing of interests already occurs in the rules governing how personal information may be handled – i.e. in the APPs themselves.

[59] Canada (Information Commissioner) v. Canada (Transportation Accident Investigation and Safety Board), 2006 FCA 157; available at https://www.canlii.org/en/ca/fca/doc/2006/2006fca157/2006fca157.html

[60] CCPA section 1798.140(o)(1)

This debate already occurred in the UK and the EU in the lead up to the drafting of the GDPR. A 2003 Court of Appeal case in the UK, *Durant v Financial Services Authority*, had read restrictively the phrase 'relate to', choosing a narrower concept of being the 'concern' or primary *subject* of the data, compared with a broader notion of being connected to the data. That decision was criticised as contrary to decisions of the higher European Court of Justice interpreting the then Data Protection Directive.

In 2006 the UK's ICO released guidelines in response, suggesting that when in doubt, the test 'relate to' ought to be read protectively; i.e. that if the data could have an adverse impact on the individual, it should be considered personal information.[61] However this heavily context-dependent approach suffers from the data holder's relative ignorance about the context affecting any particular individual within a dataset.

More recent guidance from the UK ICO,[62] issued with respect to the wording of the GDPR, states that information will "relate to" an individual if:

- the individual is named in the information, or
- the individual is the subject matter of the information (i.e. "the content is obviously about" them or their activities), or
- the information is linked to the individual (e.g. a salary is linked to a job title which is known to be held by that individual), or
- even if the person is not named, "if you are processing it to learn or record something about that individual, or where the processing has an impact on that individual", or
- if the information "is used, or is likely to be used, to learn, evaluate, treat in a certain way, make a decision about, or influence the status or behaviour of an individual", regardless of the original intent of collection.

The Macquarie Dictionary provides the following definitions:[63]

> **about**
> 1. of; concerning; in regard to: *to talk about secrets.*
> 2. connected with: *instructions about the work.*
>
> **relate**
> … 2. to bring into or establish association, connection, or relation.
>
> **concerning**
> 1. relating to; regarding; about.

[61] Mark Burdon and Paul Telford, "The Conceptual Basis of Personal Information in Australian Privacy Law", *eLaw Journal: Murdoch University Electronic Journal of Law*, 2010, Vol 17(1); available at https://eprints.qut.edu.au/37696/

[62] See https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-the-meaning-of-relates-to/

[63] *Macquarie Dictionary Online*, 2019 edition; available at https://www.macquariedictionary.com.au/

In order to resolve some of the challenges noted above, and to aim for consistency with the GDPR and many other international privacy laws, this Research Paper suggests that the definition in the Privacy Act ought to be amended to use the phrase "which is about or relates to", instead of simply "about".

A drafting note or the Explanatory Memorandum to the Bill could further explicate that 'about or relates to' means:

(i) if the individual is a subject of the information, or

(ii) if the information concerns or links to the individual, or

(iii) if the intent or effect of the information's handling will be to learn, evaluate, treat in a certain way, make a decision about, influence the status or behaviour of, or otherwise have an impact upon, the individual.

'an individual'

The current definition only covers living human beings.

Most of the State and Territory privacy laws in Australia also include deceased persons.[64] Other international statutes surveyed did not appear to include the deceased.

Only the CCPA explicitly included households and devices.

The extent to which the definition should be expanded to cover deceased individuals, groups of individuals such as households, and/or devices, is considered further below.

'who is identifiable'

The Australian Privacy Foundation, in its submission to the Australian Government in response to the DPI, stated that:

> "the definition of 'personal information' in the Privacy Act ought to be amended to clarify that it encompasses data drawn from the profiling or tracking of behaviours or movements such that an individual can be singled out (i.e. disambiguated from a crowd or cohort) and thus can be subjected to targeting or intervention, even if an

[64] For example the NSW definition of personal information provides that personal information does not include "information about an individual who has been dead for more than 30 years"; s.4(3) of the *Privacy and Personal Information Protection Act 1998* (NSW). The Victorian *Health Records Act 2001* (Vic) also sets a limit of 30 years; the limit is set at 25 years in Tasmania, five years in the Northern Territory, and "as far as is practical" in the ACT's *Health Records (Privacy and Access) Act 1997*.

> individual cannot be 'identified', in the conventional sense, from the data or related data".[65]

This begs the question: what does 'identifiable' actually mean?

The OAIC advises that:

> "Generally speaking, an individual is 'identified' when, within a group of persons, he or she is 'distinguished' from all other members of a group."[66]
>
> "The key factor to consider is whether the information can be linked back to the specific person that it relates to."[67]

The GDPR states that:

> "an identifiable person is one who can be identified, **directly or indirectly**, in particular by reference to an identifier such as a name, an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person" (emphasis added).[68]

This definition is given further context in the Recitals:

> "To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as **singling out**, either by the controller or by another person to identify the natural person directly or indirectly" (emphasis added).[69]

Guidance from the UK ICO suggests that the concept of identifiability includes online identifiers which "may be used to **distinguish** one user from another" (emphasis added).[70]

In the United States, the phrase commonly used is "personally identifiable information", or 'PII'. The definition used by the National Institute of Standards and Technology is:

> "any information about an individual maintained by an agency, including

[65] Australian Privacy Foundation, *Submission to Government on ACCC Digital Platforms Inquiry*, 10 September 2019; available at https://consult.treasury.gov.au/structural-reform-division/digital-platforms-inquiry/consultation/download_public_attachment?sqId=question-2019-07-30-0759234093-publishablefilesubquestion&uuId=848818868

[66] Office of the Australian Information Commissioner, *What is personal information?*, May 2017, p.8; available at https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information

[67] Office of the Australian Information Commissioner, *What is personal information?*, May 2017, p.8; available at https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information

[68] Article 4 of the GDPR.

[69] Recital 26 of the GDPR.

[70] See https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-are-identifiers-and-related-factors/#pd1

> (1) any information that can be used to **distinguish or trace an individual's identity**, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and
>
> (2) any other information that is **linked or linkable to an individual**, such as medical, educational, financial, and employment information" (emphasis added).[71]

A number of recent statutes and other international instruments, developed since the GDPR, have shifted beyond identifiability as a threshold requirement.

The 2018 California Consumer Privacy Act (CCPA) expressly includes, within its definition of personal information, data which is "**capable of being associated with**, or could reasonably be **linked**, directly or indirectly, with a particular consumer or household", without first needing to pass an identifiability test.[72] This theme is further fleshed out within the definition of 'unique identifier', which means "a persistent identifier that can be **used to recognize** a consumer, a family, or a device that is linked to a consumer or family, over time and across different services, including, but not limited to, a device identifier".[73]

The 2019 Nigerian privacy statute explicitly defines 'identifiable information' to include "information that can be used on its own or with other information **to identify, contact, or locate** a single person, or to identify an individual in a context" within its definition of 'personal data' (emphasis added).[74]

The 2019 international standard in Privacy Information Management, ISO 27701, uses the phrase:

> "information that
>
> (a) can be used to identify the PII principal to whom such information relates, or
>
> (b) is or might be **directly or indirectly linked to** a PII principal" (emphasis added)

within its definition of 'personally identifiable information'.[75]

Each of these recent developments has either explicitly broadened the notion of identifiability, or has introduced *alternatives* to identifiability as a threshold element of the definition.

[71] National Institute of Standards and Technology, Special Publication 800-122: *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*, April 2010, p.ES-1; available at https://www.nist.gov/publications/guide-protecting-confidentiality-personally-identifiable-information-pii

[72] CCPA section 1798.140(o)(1)

[73] CCPA section 1798.140(o)(1)(x)

[74] *Nigeria Data Protection Regulation 2019*; available at https://nitda.gov.ng/wp-content/uploads/2019/01/NigeriaDataProtectionRegulation.pdf

[75] ISO/IEC 27701:2019, available at https://www.iso.org/standard/71670.html

The Macquarie Dictionary provides the following definitions:[76]

> **identity:**
> …
> 4. condition, character, or distinguishing features of person or things: *a case of mistaken identity.*
> 5. official information about yourself, as passports, bank account details, etc.; ID: *keep your identity safe.*
>
> **identify:**
> 1. to recognise or establish as being a particular person or thing; attest or prove to be as claimed or asserted: *to identify handwriting*; *to identify the bearer of a cheque.*
> 2. to serve as a means of identification for: *this card identifies the bearer as a member.*
>
> **identifiable:**
> 1. able to be identified.
> 2. discernible; recognisable
>
> **discern**
> 1. to perceive by the sight or some other sense or by the intellect; see, recognise, or apprehend clearly.
> 2. to distinguish mentally; recognise as distinct or different; discriminate: *he discerns good and bad*; *he discerns good from bad.*
> 3. to distinguish or discriminate.
>
> **individuation**
> 1. the act of individuating.
> 2. the state of being individuated; individual existence; individuality.
> 3. *Philosophy* the determination or contraction of a general nature to an individual mode of existence; development of the individual from the general.
>
> **individuate:**
> 1. to form an individual or distinct entity.
> 2. to give an individual or distinctive character to; individualise.

Even the word ‘identity’ has multiple meanings, only one of which is identity in the sense of something officially designated and verifiable; used in the sense of “Show me your ID”, it encompasses legal name, or perhaps a combination of legal name, date of birth, and government-issued unique identifier. Another is simply the “condition, character, or distinguishing features of person”, which does not necessarily relate to a verifiable or official claim.

[76] *Macquarie Dictionary Online*, 2019 edition; available at https://www.macquariedictionary.com.au/

‘Identifiable’ can mean “able to be identified”, but also “discernible” or “recognisable”. To ‘discern’ is to “recognise as distinct or different”.

Professor Bygrave argues that, at least under European data protection law with its inclusion of ‘indirect’ identifiability in its statutes since the 1990s:

> “there is little doubt that the ability to identify a person is essentially the ability to distinguish that person from others by linking him/her to pre-collected information of some kind. As such, identification does not require knowledge of a person's name but it does require knowledge of some unique characteristics of the person relative to a set of other persons”.[77]

As noted above, newer statutes from California and Nigeria, and a new international standard, as well as guidance from NIST, each suggest that they see data which is ‘linkable’ to an individual as within the scope of privacy protection, *regardless* of whether the individual can be identified.

The extent to which Australian privacy law does or should incorporate the concept of ‘linkability’ or ‘singling out’ within the notion of ‘identifiable’ is discussed further below; see *Individuation*.

Identifiable in isolation?

Professor Lee Bygrave poses this as the question: “to what extent is the use of auxiliary information permitted in the identification process?”[78]

Across many jurisdictions, the legal definitions of personal information or personal data typically allow for some combination of data when testing for identifiability, such that considering the identifiability of a dataset in isolation is not enough.

For example, GDPR Recital 26 refers to “taking into consideration the available technology”, which means that in determining whether a person is identifiable, account should be taken of all the means reasonably likely to be used to enable identification or re-identification.

It is now relatively settled in Australian jurisprudence that identifiability is not to be determined by examining *only* the data at issue.

In 2014, the definition of ‘personal information’ in the Privacy Act was revised. It had previously referred to “…about an individual whose identity is apparent, or can reasonably be ascertained, *from the information or opinion*. Removing “from the…” clarified that the information need not *prima facie* lead to identification; it can be considered in conjunction with other sources.

[77] Lee Bygrave, *Data protection law: approaching its rationale, logic and limits*, Kluwer Law International, 2002, part 10.4.
[78] Lee Bygrave, *Data protection law: approaching its rationale, logic and limits*, Kluwer Law International, 2002, part 2.4.1.

The OAIC has advised:

> "Under the Privacy Act, an individual will be 'identifiable' where it is possible to identify the individual from available information, including, but not limited to, the information in issue".[79]
>
> "Some information may not be personal information when considered on its own. However, when combined with other information held by (or accessible to) an entity, it may become 'personal information'. Information holdings can therefore be dynamic, and the character of information can change over time".[80]

NCAT has found that if a vehicle registration number is collected in isolation from any other data, and is not recorded in a meaningful way such that it could be linked back to any other identifying information, then in isolation the vehicle registration number will not be 'personal information'. However NCAT also noted that when held in a different system, along with other details such as name of the vehicle's owner or driver, the vehicle registration number is clearly personal information. The critical detail in that case was that a parking meter briefly transmitted the vehicle registration number to a server, in order to check whether that number was held on a list of numbers for which a parking-payment exemption had been granted. Neither the meter nor the server held a record that the verification check had occurred, and thus the collection of the vehicle registration number was only transitory. In the absence of a record, there was no pragmatic way for a link to be made between the vehicle registration number entered at the meter, and the separate database held by council which included identifying details.[81]

Other jurisdictions in which regulatory guidance or caselaw allows identifiability to be judged with reference to additional materials include Victoria[82] and Canada.[83]

How easily identifiable?

The Australian Privacy Act's definition of personal information includes the qualifier 'reasonably' ahead of 'identifiable'. All other Australian State and Territory privacy laws also have a qualifier, though it is expressed differently between the jurisdictions, such as the

[79] Office of the Australian Information Commissioner, *Publication of MBS/PBS data: Commissioner initiated investigation report*, 23 March 2018, p.4; available at https://www.oaic.gov.au/assets/privacy/privacy-decisions/investigation-reports/publication-of-mbs-pbs-data.pdf

[80] Office of the Australian Information Commissioner, *What is personal information?*, May 2017, p.6; available at https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information

[81] *DAB v Byron Shire Council* [2017] NSWCATAD 104; available at https://www.caselaw.nsw.gov.au/decision/58dd93a7e4b0e71e17f58564; see also *WL v Randwick City Council (No. 2)* [2010] NSWADT 84, *APV and APW and Department of Finance and Services* [2014] NSWCATAD 10, and *Office of Finance and Services v APV and APW* [2014] NSWCATAP 88.

[82] *WL v La Trobe University (General)* [2005] VCAT 2592; available at http://www7.austlii.edu.au/cgi-bin/viewdoc/au/cases/vic/VCAT/2005/2592.html

[83] *Gordon v. Canada (Health)*, 2008 FC 258 at [34]; available at https://www.canlii.org/en/ca/fct/doc/2008/2008fc258/2008fc258.html

phrase “whose identity is apparent, or can reasonably be ascertained”,[84] “reasonably ascertainable”,[85] or “reasonably identifiable”[86].

However internationally, no other laws that we surveyed for this Research Paper used any qualifier to limit the notion of identifiability. The GDPR; the national privacy laws of New Zealand, Canada, Singapore, South Africa, Brazil, Nigeria, Japan, Hong Kong and India; the Council of Europe’s Convention 108 and 108+; and the 1980 *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* (including the 2013 update) all allow for ‘identifiability’ to be considered without needing qualification.

The OAIC’s guidance is that: “An individual will be ‘reasonably’ identifiable where the process or steps for that individual to be identifiable are reasonable to achieve”.[87]

> “whether the process of identification is reasonable to achieve … is determined by asking whether, objectively speaking, it is reasonable to expect that the subject of the information could be identified. Even though it may be technically possible to identify an individual from information, if doing so is so impractical that there is almost no likelihood of it occurring, the information would not generally be regarded as ‘personal information’.
>
> Determining whether a person is ‘reasonably’ identifiable will require a contextual consideration… including:
>
> a) the nature and amount of information
>
> b) who will hold and have access to the information, and
>
> c) the other information that is available, and the practicability of using that information to identify an individual.” [88]

Therefore the context in which the data is held or released, and the availability of other datasets or other resources to attempt a linkage, will be key to determining whether any individual is identifiable:

> “Whether an individual is reasonably identifiable depends on the nature of the information in issue, and the context in which the information is held or released”.[89]
>
> For example, “By itself, information that allows Nina to be contacted — such as a telephone number or a street address — may not be about an ‘identifiable’ individual.

[84] *Privacy and Personal Information Protection Act 1998* (NSW), s.4; *Information Privacy Act 2009* (Qld), s.12; *Privacy and Data Protection Act 2014* (Vic) s.3.
[85] *Information Act 2002* (NT), s.4A; *Personal Information Protection Act 2004* (Tas), s.3.
[86] *Information Privacy Act* (ACT), s.8.
[87] Office of the Australian Information Commissioner, *Publication of MBS/PBS data: Commissioner initiated investigation report*, 23 March 2018, p.4; available at https://www.oaic.gov.au/assets/privacy/privacy-decisions/investigation-reports/publication-of-mbs-pbs-data.pdf
[88] Office of the Australian Information Commissioner, *What is personal information?*, May 2017, p.8; available at https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information
[89] Office of the Australian Information Commissioner, *Publication of MBS/PBS data: Commissioner initiated investigation report*, 23 March 2018, p.3; available at https://www.oaic.gov.au/assets/privacy/privacy-decisions/investigation-reports/publication-of-mbs-pbs-data.pdf

> However, Nina is likely to be identified where this information can be used to search a business's customer database, locating an entry about Nina".[90]

The National Health and Medical Research Council (NHMRC) likewise warns researchers that identifiability can be affected by contextual factors, including who has access to the data, and their technical capabilities to enable re-identification.[91]

Who will hold and have access to the information is therefore a relevant consideration when assessing whether an individual will be 'reasonably identifiable'. In relation to information publicly released, the OAIC makes the point that it is "difficult to anticipate who might access the information, what other types of information they have access to for referencing purposes, and what motivations they may have to identify an individual".[92]

Identifiable by whom?

A question arises as to *who* is supposed to be able to ascertain a person's identity – is it the subject person themselves, the person collecting or using the information in question, a wider audience, or indeed a machine?

Professor Lee Bygrave poses this as the question: "who is the legally relevant agent of identification (ie, the person who is to carry out identification)?"[93]

Australian jurisprudence is now relatively settled that this test is to be applied broadly, but will nonetheless be informed by context; in other words, the test is *anyone* (or possibly even a machine) who might *receive or have access to* the information.

The OAIC provides the following example: "Mark's licence plate number may be 'personal information' when held or disclosed to an employee of a law enforcement agency, but may not be personal information in the hands of an average member of the public".[94]

Guidance from the NSW Privacy Commissioner states that "The test is whether identification is possible, by any person (or machine) other than the subject themselves".[95]

[90] Office of the Australian Information Commissioner, *What is personal information?*, May 2017, p.8; available at https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information

[91] National Health and Medical Research Council, Australian Research Council and Universities Australia, *National Statement on Ethical Conduct in Human Research, 2007 (2018 update)*, p.33; available at https://nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018.

[92] Office of the Australian Information Commissioner, *What is personal information?*, May 2017, p.9; available at https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information

[93] Lee Bygrave, *Data protection law: approaching its rationale, logic and limits*, Kluwer Law International, 2002, part 2.4.1.

[94] Office of the Australian Information Commissioner, *What is personal information?*, May 2017, p.9; available at https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information

[95] NSW Information & Privacy Commission, *Fact Sheet: Reasonably Ascertainable Identity*, January 2017, FS 2017/001; available at https://www.ipc.nsw.gov.au/fact-sheet-reasonably-ascertainable-identity-0

The NSW Privacy Commissioner has previously found that information which identified a student in his "school community" and "local community" amounted to "personal information", notwithstanding that the discloser did not know the student's identity.[96]

In that case, enough information had been mentioned on air by the Education Minister for a particular audience – the "school community" – to be able to ascertain the student's identity. Although he did not name the student or the school in question, the Minister had identified the student's gender, age, the year in which he was enrolled, a description of an event involving the student, the date on which the student was removed from the school, and the date on which a school assembly was planned.

The NSW Privacy Commissioner noted that: "The question at issue is whether the combination of items of information about Student A reasonably enabled the receivers of the information to ascertain or deduce Student A's identity".[97] The Commissioner's investigation concluded that it did. Indeed he found that from the information that was disclosed on air by the Minister, not only could the local and school community identify the student and his family (in a way that had not been possible before the Minister's comments on air), but within a day the wider media had identified and published the name of the school and the student in question.

Thus it is possible that even if an individual or organisation collecting, holding or using some information does not know the subject person's identity, and could not reasonably ascertain their identity from the information in their possession, they may nonetheless be handling "personal information", because another audience could make that link.

Similarly, in a case involving CCTV footage from a shopping centre, NCAT found that footage of an assault contained not only the personal information of the offender and the victim, who were already known to the police, but also the personal information of bystanders, only some of whom had been identified by the police:

> "whether the police had been unable to identify them or not is irrelevant to whether it contains their personal information. Some members of the public would have had access to publicly available information about the identities of those other persons, for example through social media, and consequently the footage contained personal information about them".[98]

[96] Privacy NSW, Special Report to Parliament, *Student A and the Minister for Education*, 7 May 2002.
[97] Privacy NSW, Special Report to Parliament, *Student A and the Minister for Education*, 7 May 2002.
[98] *Field v Commissioner of Police, New South Wales Police Force* [2015] NSWCATAD 153 at [33]; available at https://www.caselaw.nsw.gov.au/decision/55ac5080e4b06e6e9f0f7c69; see also *Meldru v Wollondilly Shire Council* [2017] NSWCATAD 292; available at https://www.caselaw.nsw.gov.au/decision/59cb5245e4b058596cbaab7a

Indeed potentially, identification can be performed by *no* individual, if it is performed by a *machine* instead of a human. In a NSW case involving redacted documents, the accepted facts were that the method used to 'redact' AIN's name from a document prior to its publication on the respondent's website was insufficient, because it:

> "merely placed an opaque 'block' over the top of the 'redacted' words. While this prevented the human eye from reading the Applicant's name, it allowed, as was later discovered, Google webcrawlers (also called "Googlebots") to read this information and to link the publication to a search of 'Dr [AIN]'".[99]

As a result, a Google search against the applicant's name yielded a link to a copy of the redacted document. NCAT found that the applicant's personal information had been disclosed, noting that:

> "While the Applicant's personal information was masked from the human eye, her personal information was able to be 'read' by the Google search engine. This resulted in a search for 'Dr [AIN]' (or similar) leading to a link to a copy of the (human eye redacted) Medical Tribunal's decision. The Respondent accepted, properly in my view, that a Google search for 'Dr [AIN]' would link the Medical Tribunal's decision to her" (at [32]).[100]

Other jurisdictions have taken a variety of approaches. A Hong Kong case suggested that the subject person's identity must be apparent to, or reasonably ascertainable by, the "data user" – that is, the individual or organisation collecting, holding or using the information.[101] One test developed in New Zealand in the context of media disclosures requires "that the person be identifiable beyond their intimate friends or family", while the New Zealand Privacy Commissioner prefers the broader test "if the person can be recognised by others than themselves".[102]

However a complaint about a photograph was dismissed by the New Zealand Privacy Commissioner because the complainant was the only person who could possibly recognise himself from a photograph – only the complainant knew that he was in the frame at the time, and the photograph did not show enough detail for any other viewer to even be able to determine the person's gender, let alone more detail such as to guess at the person's identity.[103]

[99] *AIN v Medical Council of New South Wales* [2016] NSWCATAD 5 at [59]; available at https://www.caselaw.nsw.gov.au/decision/5689d138e4b05f2c4f04a6a8

[100] The Tribunal's findings of fact were overturned on appeal, when the Appeal Panel found that in fact the 'redacted' information could also be read by the human eye, because any person with a PDF editing tool was able to remove the 'blanking' to display AIN's name; *AIN v Medical Council of New South Wales* [2017] NSWCATAP 23. However this did not affect the earlier conclusion about the machine-readable text.

[101] Eastweek Publisher Ltd. and Another v Privacy Commissioner for Personal Data [2000] HKCA 137.

[102] Katrine Evans, "Personal information in New Zealand: Between a rock and a hard place?", paper for *Interpreting Privacy Principles: Chaos or Consistency*?, Symposium, Sydney, 17 May 2006.

[103] Katrine Evans, "Personal information in New Zealand: Between a rock and a hard place?", paper for *Interpreting Privacy Principles: Chaos or Consistency*?, Symposium, Sydney, 17 May 2006.

De-identified data

The flip side of the definition of ‘personal information’ is what is currently known as ‘de-identified’ information.

Globally, information privacy or ‘data protection’ laws have as their starting point some notion of *identifiability*. The commonality between all these different laws, jurisdictions and legal definitions is that if no individual is *identifiable* from a set of data, then the relevant privacy principles (or other legal obligations, however expressed) simply won’t apply.

Thus the idea of ‘de-identified’ or ‘anonymous’ data is that it represents the end-point where privacy laws will cease to apply, because no individual can be identified from the data.

Section 6 of the Privacy Act states:

> “personal information is ***de-identified*** if the information is no longer about an identifiable individual or an individual who is reasonably identifiable”.

However de-identification does not always result in truly anonymous data. The de-identification of ‘big data’ datasets is complex. The OAIC has warned that de-identification “can be effective in preventing re-identification of an individual, but may not remove that risk altogether”, for example if “another dataset or other information could be matched with the de-identified information”.[104] There is dispute within the privacy and computer science fields as to whether achieving anonymous data is even possible.[105]

There are two practical difficulties with the current definition which need resolution:

- the word ‘de-identified’ is highly ambiguous because it means different things to different audiences, and

- the phrase ‘is no longer’ precludes privacy risks arising from this dataset for third parties.

In addition, the language of “reasonably identifiable” is problematic, as already explored above in relation to the definition of ‘personal information’.

[104] Office of the Australian Information Commissioner, *De-identification of data and information*, April 2015; was previously available at https://www.oaic.gov.au/information-policy/information-policy-resources/information-policy-agency-resource-1-de-identification-of-data-and-information but has since been replaced by the more recent publication: OAIC, *De-identification and the Privacy Act*, 21 March 2018, available at https://www.oaic.gov.au/privacy/guidance-and-advice/de-identification-and-the-privacy-act/

[105] For competing views, see the work of a former Privacy Commissioner and Washington think tank (Cavoukian and Castro, “Big Data and Innovation, Setting the Record Straight: De-identification *Does* Work”, June 2014; available at http://www2.itif.org/2014-big-data-deidentification.pdf), and in direct response the work of Princeton computer science scholars (Narayanan and Felten, “No silver bullet: De-identification still doesn't work”, July 2014; available at https://www.cs.princeton.edu/~arvindn/publications/no-silver-bullet-de-identification.pdf).

This review of the definition of 'personal information' is an opportunity to also clarify the terminology used in relation to de-identification in the Privacy Act, in order to offer greater certainty for regulated entities, and aim for consistency with the GDPR.

Confusion about terminology heightens risks

Earlier in this Research Paper we highlighted the many examples of supposedly 'de-identified' datasets released publicly, which then turn out to enable identification or re-identification of individuals.

One of the causes is a disconnect between the terminology used by statisticians and that used by people concerned with applying privacy law. This disconnect leads privacy risk assessments to be concluded on incorrect assumptions, leading to unauthorised disclosures of personal information.[106]

The Office of the Victorian Information Commissioner has noted that the "word 'de-identify' is, unfortunately, highly ambiguous." [107]

Indeed the NHMRC, the Australian Research Council and Universities Australia, which jointly govern research in Australia, avoid using the word 'de-identified' at all, because of the ambiguity surrounding the term. They have noted the potential confusion arising between those who use the word 'de-identified' to describe non-identifiable data, and those who use it to describe data from which identifiers have been removed - but which could nonetheless pose a re-identification risk.[108]

Thus when a statistician or data scientist talks about *de-identified data*, they do not necessarily mean that no-one is identifiable from the data. They simply mean that de-identification techniques have been applied to the data, in an *attempt* to minimise the likelihood that any individual's identity will be, *prima facie*, revealed. The nature of the treatment might be one of a number of techniques, including suppression, generalisation, micro-aggregation, pseudonymisation, hashing, etc. The words 'confidentialisation' and 'anonymisation' are used interchangeably with 'de-identification' to mean the same thing.

However to a lawyer or privacy professional, or indeed to an affected consumer or citizen, the word 'de-identified' can be heard as a promise or guarantee that no individual could be identified from that dataset; and that therefore privacy protections and privacy rights will not apply to the data.

[106] The MBS/PBS data disclosure and the Myki dataset were each examples of this problem: Privacy Commissioner investigations showed that insufficient understanding about re-identification risks led to datasets being made public.

[107] Office of the Victorian Information Commissioner, *Protecting unit-record level personal information: The limitations of de-identification and the implications for the Privacy and Data Protection Act 2014*, May 2014, p.5; available at https://ovic.vic.gov.au/wp-content/uploads/2018/07/Protecting-unit-record-level-personal-information.pdf

[108] National Health and Medical Research Council, Australian Research Council and Universities Australia, *National Statement on Ethical Conduct in Human Research, 2007 (2015 update)*, p.27.

A typical conversation between the two parties in a privacy risk assessment process might go something like this:

> Privacy officer: “Is the data de-identified?”
>
> > *(Meaning: Have we met the legal test under the Privacy Act such that no individual could be reasonably identified, and thus the APPs won’t apply?)*
>
> Data scientist: “Yes, the data has been de-identified.”
>
> > *(Meaning: We pseudonymised all direct identifiers like names, and we hashed the event dates.)*

However, just because data has been ‘de-identified’ in a statistical sense does not make it ‘de-identified’ in the legal sense. Identifiability risks might remain, either because the statistical techniques do not withstand a motivated intruder attack (for example, pseudonyms can still be interrogated, hashing can be reversed), or because there is enough detail in the remaining data to link to some identifiable individuals (for example, a person with a rare disease will stand out in a dataset of medical records).

The Privacy Act should be amended to ensure that a description of how data has been treated in a *statistical* sense should no longer be conflated with an assertion about its *legal* status under the Privacy Act. The word ‘de-identified’, like ‘confidentialised’ or ‘anonymised’, should mean simply that one or more de-identification techniques have been applied to a dataset.

By using a word in the Privacy Act which is an adjective (‘anonymous’), instead of a word which can be either an adjective or a verb (‘de-identified’), this problem of conflation between the two meanings should be ameliorated, if not completely resolved. The term ‘anonymous’ is also consistent with the GDPR.

Indeed we counsel against using the word ‘de-identified’ at all, instead preferring the descriptor phrase “data to which de-identification techniques have been applied”, to make clear that a description of how data has been treated is neither a guarantee about the residual risk of individuals being identifiable from the data, nor a statement about whether or not the Privacy Act applies to the data.

The three types of privacy risks from 'de-identified' data

There are three types of privacy risks which can arise from a supposedly 'de-identified' dataset:

- identity disclosure,
- attribute disclosure, or
- individuation.

The Privacy Act's definition of 'de-identified' does not currently control for all three types.

Privacy harms can arise from various practices, one of which is the disclosure of previously unknown information about a person. De-identification aims to break the link between a dataset and an individual in the real world, so that the disclosure of a fact (such as that a patient is being treated at this hospital for HIV) cannot be linked back to an identified individual (the patient is Sally Citizen).

The harm being prevented here is known as 'identity disclosure'. Identity disclosure - which occurs when data is re-identified - can arise by one of two ways: by either matching a person to data, or matching data to a person.

A re-identification attempt might involve taking an individual, and finding data that matches them. You choose an individual about whom you know certain facts - you likely know the date of birth and home address of your friends and family, or from newspaper reports you might know about a celebrity who has just had a knee operation - and see if you can find details to match them within the dataset. For example: you know Sally Citizen is a patient at this hospital, but can you find out her health information? A 'linkage attack' will use data to identify people by linking the record with other information known about them.

Alternatively, a re-identification attempt can involve starting with the data, and finding the individual to whom that data relates. For example: you have the clinical records for all patients at this hospital, but can you find out who one or more of those records relate to? A 'cryptographic attack' could for example seek to reverse the encryption algorithm used to hide patient names or numbers within the dataset. Normally encryption algorithms are extraordinarily difficult to reverse, but in special circumstances, such as where the encrypted data is known to have a special fixed format, reversal becomes feasible.

Examples where individuals were able to be re-identified from publicly released data include:

- New York City taxi trip data (data-to-person re-identification of all taxi drivers); see *Case study: NYC taxi trip data* below

- medical practitioner details which could be recovered from MBS/PBS records publicly released by the Australian Government (data-to-person re-identification of all health service providers)[109], and

- the re-identification of patient details from the same MBS/PBS dataset, using other known facts about individuals (person-to-data re-identification of some patients).[110]

A person can also suffer harm through 'attribute disclosure', which refers to when you learn new facts about an already-identified individual. The release of New York City taxi trip data provides an example. (See *Case study: NYC taxi trip data* below.) It was already 'known' (from other sources) that celebrities got into taxis at certain points. However from the disclosure of the taxi trip data, more information could then be established about those celebrities, such as where they went and how much they paid.

Similarly, a member of the public could interrogate the data to find out where a known person (say, a spouse), who was known to have arrived at a particular location at a particular date and time (say, home at 1am on Tuesday 3 March), had *begun* their taxi trip; did the taxi trip originate at their office (as the spouse may have claimed), or somewhere else? The interrogator may draw some inferences depending on the result of their query.

The third way in which a person might suffer harm is if a third party could draw inferences about a *stranger*, and target them for harm in some way, without necessarily knowing their *identity*. Again using the taxi trip data as an example, a motivated person could look up the end point for all taxi trips that started at an address known to be a strip club, or an abortion clinic, or an Alcoholics Anonymous meeting venue, and draw inferences - rightly or wrongly - about the people who live at the addresses where those taxi trips ended.[111] Individuals could be targeted for intervention or harm based on data, without the data user ever knowing who they are. This type of privacy harm arises from *individuation*, even if there is no *identifiability* component.

The current definitions of 'personal information' and 'de-identified' in the Privacy Act fail to consider the privacy risks posed by individuation. Our recommendations suggest a re-think of the definitions, to resolve this problem.

[109] Dr Chris Culnane, Dr Benjamin Rubinstein and Dr Vanessa Teague, "Understanding the maths is crucial for protecting privacy", University of Melbourne blog, 29 September 2016; at https://pursuit.unimelb.edu.au/articles/understanding-the-maths-is-crucial-for-protecting-privacy

[110] Office of the Victorian Information Commissioner, *Protecting unit-record level personal information: The limitations of de-identification and the implications for the Privacy and Data Protection Act 2014*, May 2018, p.16; available at https://ovic.vic.gov.au/wp-content/uploads/2018/07/Protecting-unit-record-level-personal-information.pdf

[111] The definition of 'personal information' in Australia explicitly includes information about an individual, whether or not it is correct.

The risk to third parties

If de-identification practices are focussed solely on treating the ostensible ‘data subjects’ in a dataset, any risk analysis (such as testing for the possibility of *re*-identification) will fail to consider the risks arising for third parties.

The taxi trip dataset offers a perfect example. The only ‘subjects’ whose personal information was ever collected by the NYC Taxi and Limousine Commission was that of taxi drivers. The Commission itself held no data about taxi *passengers*. Yet data about the taxi trips, once combined with other data sources, was capable of revealing personal information about passengers.

The language of de-identification typically talks about de-identifying data, and testing for *re*-identification risks. For example, the current definition of ‘de-identified’ uses the language of ‘is no longer about…’, which pre-supposes that the only people whose privacy is to be protected via de-identification techniques *were* identifiable in the dataset to start with. This precludes privacy risks arising from any given dataset for third parties who were never the original ‘subjects’, but who can nonetheless suffer privacy harms from inappropriate use or disclosure of the dataset.

The current definition of ‘de-identified’ in the Privacy Act fails to consider the privacy risks posed to third parties. Our recommendations suggest a re-think of the definition, to resolve this problem.

Legislative solutions

If the definition of ‘personal information’ in the Privacy Act is to be amended, so too must the definition of ‘de-identified’.

One health privacy law in the United States, the *Health Insurance Portability and Accountability Act* of 1996 (HIPAA), creates an ‘anonymization standard’: the HIPAA Privacy Rule holds that data has been de-identified if 18 types of identifiers have been removed, or if an expert has determined the risk of re-identification to be ‘very small’.[112]

However this approach to managing the privacy risks arising from datasets contains several weaknesses, including that it fails to anticipate that the residual attribute data might still enable re-identification (e.g. patients with rare diseases), or the residual data might be linkable to other data sources. It does not anticipate any risks posed to third parties, nor the risks arising from individuation.

[112] The 18 elements are listed at https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/

Furthermore, by embedding the list of elements in statute, it quickly becomes out of date: the list includes fax numbers, but not for example a username or login ID for a health app. This offers a lesson in the risks posed by being too prescriptive or calling out particular technologies within the wording of statute law.

While there is some merit in adopting the full definition of 'anonymous data' from the GDPR, the GDPR definition also fails to encompass the risks posed to third parties. Article 4 of the GDPR uses the phrase "rendered anonymous in such a manner that the data subject is not or no longer identifiable". By focussing the test on only the 'data subject', data meeting this definition could still potentially reveal information about a third party (e.g. the passenger of a taxi in the taxi trip dataset scenario, because only the taxi *driver* would be the 'data subject').

A more elegant solution, affording simplicity of language, legal clarity and the protection of third parties, would be to define "anonymous data" as "data from which no individual is identifiable".

This definition would offer:

- global consistency by matching the intent of the GDPR
- internal consistency with the proposed other amendments to the definitions of 'personal information' and 'identifiable'
- greater clarity by distinguishing between the legal concept of 'anonymous data' and the statistical techniques of 'de-identification', and
- improved protection against privacy harms involving either attribute disclosure of third parties, or individuation.

Individuation

The assumption upon which data privacy laws rest is that identifiability is the key to harm: "the underlying conceptual focus of defining personal information in Australian privacy laws regards the revealment of identity as the social harm to be protected".[113]

In other words, the assumption is that no harm can befall an individual from the handling of their personal data if they cannot be *identified* from the data; that information which might otherwise cause embarrassment, humiliation, or physical, psychological or financial risks cannot cause such harms if no-one knows *who* the information is about.

However in the 21st century, that assumption is no longer true.

From the digital breadcrumbs we leave behind in the form of geolocation data shed from our mobile devices, to the patterns of behaviour we exhibit online as we browse, click, comment, shop, share and 'like', we can be tracked. Tracked, traced, monitored, surveilled; then profiled; and finally targeted … all without the party doing the tracking, profiling or targeting needing to know 'who' we are.

The digital environment has turned on its head the assumption that identifiability – in the sense of knowing a person's 'identity' - is only vector for privacy harm. As the OAIC has noted, "harm can be caused by just knowing attributes of an individual, without knowing their identity".[114]

The most obvious – and disturbing – recent example is the finding that publicly disclosed data about public transport cards used in Melbourne, though purportedly 'de-identified', could have been used to determine where patterns could be found showing young children travelling without an accompanying adult. Those children could be targeted by a violent predator as a result, without the perpetrator needing to know anything about the child's identity. (See *Case study: Myki public transport data*, below.)

Other examples of potential harm arising even without identities being revealed have included the release of data about taxi trips and consumers' fitness routines; see *Case study: NYC taxi trip data*, and *Case study: Strava fitness data*, below.

If the objective of privacy laws is to protect people's privacy, those laws need to grapple with a broader view of the types of practices which can harm privacy – regardless of whether an individual's identity is known or revealed.

[113] Mark Burdon and Paul Telford, "The Conceptual Basis of Personal Information in Australian Privacy Law", *eLaw Journal: Murdoch University Electronic Journal of Law*, 2010, Vol 17(1), p.27; available at https://eprints.qut.edu.au/37696/

[114] Office of the Australian Information Commissioner, *What is personal information?*, May 2017, p.21; available at https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information

This Research Paper uses the word *individuation* to refer to the ability to disambiguate or 'single out' a person in the crowd, even if that individual's 'identity' is not known.[115]

Individuation is the technique used in online behavioural advertising; advertisers don't know who any particular consumer is, but they know that the user of a particular device has a certain collection of attributes, and they can target or address their message to the user of this device accordingly.

The objective of online behavioural advertising is, like any advertising, to predict purchasing interests, and drive purchasing decisions. Online, however, the repercussions are much greater, because of the degree to which advertising – and indeed, the very content users are shown – has become 'personalised'. Personalisation means decisions are made about who sees what, and equally what will be withheld from whom.

By allowing exclusion, digital platforms also allow discrimination. Facebook has been caught allowing advertisers to target – and exclude – people on the basis of their 'racial affinity', amongst other social, demographic, racial and religious characteristics.[116] For example, a landlord with an advertisement for rental housing could prevent people profiled as 'single mothers' from ever seeing their ad; an employer could prevent people identifying as Jewish from seeing a job ad; or a bank could prevent people categorised as 'liking African American content' from seeing an ad for a home loan.[117]

Existing patterns of social exclusion, economic inequality, prejudice and discrimination are further entrenched by micro-targeted advertising, which is hidden from public view and regulatory scrutiny. Preying on vulnerable individuals which could lead to physical, financial or social harm is also a risk of micro-targeting. It was revealed in 2017 that Australian Facebook executives were promoting to advertisers their ability to target psychologically vulnerable teenagers.[118] Advertising mental health services is one thing; advertising pharmaceuticals is another; while advertising services such as high-stakes gambling to vulnerable individuals inferred by a digital platform to be in the midst of a manic episode is yet another.

'Personalisation' can lead to price discrimination, like pricing based on an airline knowing this user has searched for a quote before; or market exclusion, like insurance products only being advertised to users already profiled as 'low risk', based on their online activities.[119] Micro-targeting can also be used to manipulate behaviour, such as voting intentions.[120]

[115] Anna Johnston, "Individuation – Re-thinking the scope of privacy laws", 30 August 2016, Salinger Privacy blog; available at https://www.salingerprivacy.com.au/2016/08/30/individuation/

[116] Julia Angwin, Ariana Tobin and Madeleine Varner, "Facebook (Still) Letting Housing Advertisers Exclude Users by Race", *ProPublica*, 17 November 2017; available at https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin

[117] Alex Hern, "Facebook lets advertisers target users based on sensitive interests", *The Guardian*, 16 May 2018; available at https://amp.theguardian.com/technology/2018/may/16/facebook-lets-advertisers-target-users-based-on-sensitive-interests

[118] Nitasha Tiku, "Get Ready for the Next Big Privacy Backlash Against Facebook", *Wired*, 21 May 2017; available at https://www.wired.com/2017/05/welcome-next-phase-facebook-backlash/

[119] Rafi Mohammed, "How Retailers Use Personalized Prices to Test What You're Willing to Pay", *Harvard Business Review*, 20 October 2017; available at https://hbr.org/2017/10/how-retailers-use-personalized-prices-to-test-what-youre-willing-to-pay

[120] Luke Dormehl, "Will Your Computer Tell You How to Vote?", *Politico Magazine*, 25 November 2014; available at https://www.politico.com/magazine/story/2014/11/computers-algorithms-tell-you-how-to-vote-113142

The Facebook / Cornell University research project on emotional contagion, revealed in 2014, offers another fine example of causing privacy harm, without 'personal information' being involved. The project deliberately manipulated the news feeds of almost 700,000 unsuspecting Facebook users, and monitored their reactions, in order to trigger emotional outcomes for people who had no idea they were even part of a 'research' project. The researchers described their project as "a massive (*N* = 689,003) experiment on Facebook", which demonstrated "that emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness".[121]

The university researchers argued that because they did not know who their research subjects were, there was no identifiable personal data at stake. On that basis, Cornell University's Institutional Review Board had concluded that the academics were "not directly engaged in human research", and therefore no ethical review was required.[122] Clearly users did not see the issue this way, with typical reactions being descriptions such as 'creepy', 'evil, 'terrifying' and 'super disturbing'.[123]

These harms are privacy harms

The UN's Special Rapporteur on Privacy, Joe Cannataci, has written about privacy as enabling the free, unhindered development of personality.[124] You could think of privacy as related to the right to self-determination, or as an element of autonomy.

The activities described above hold the potential to impact on individuals' autonomy, by narrowing or altering their market or life choices. Philosophy professor Michael Lynch has said that "taking you out of the decision-making equation" matters because "autonomy enables us to shape our own decisions and make ones that are in line with our deepest preferences and convictions. Autonomy lies at the heart of our humanity".[125]

A person does not need to be identified in order for their autonomy to be undermined or their dignity to be damaged.

Much effort is expended by advertisers and others wishing to track people's movements and behaviours, whether offline or online, in convincing privacy regulators and consumers that their data is not identifying, and that therefore there is no cause for alarm. Whether in double-blind data matching models or the use of homomorphic encryption to compare data from multiple different sources (the sharing of which would be prohibited if the data was

[121] Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks", *Proceedings of the National Academy of Sciences of the United States of America*, 17 June 2014; available at https://www.pnas.org/content/111/24/8788.full

[122] Robinson Meyer, "Everything We Know About Facebook's Secret Mood Manipulation Experiment", *The Atlantic*, 28 June 2014; available at https://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/

[123] "Facebook probed on mood experiment", *SBS News*, 2 July 2014; available at https://www.sbs.com.au/news/facebook-probed-on-mood-experiment

[124] Joe Cannataci, "Privacy, personality and flows of information – an open invitation", personal blog, 9 June 2016; available at https://www.privacyandpersonality.org/2016/06/privacy-personality-and-flows-of-information-an-open-invitation/

[125] Michael Lynch, "Why does our privacy really matter?", *Christian Science Monitor*, 22 April 2016; available at https://www.csmonitor.com/World/Passcode/Security-culture/2016/0422/Why-does-our-privacy-really-matter

‘identifiable’), the current obsession is how to avoid *identifying* anybody, such that the activity can proceed unregulated by the Privacy Act. In fact the real question both companies and governments should be asking is how to avoid *harming* anybody.

Bruce Schneier has argued that laws concerned with identifiability as key element are too limiting in their treatment of potential harm:

> “most of the time, it doesn’t matter if identification isn’t tied to a real name. What’s important is that we can be consistently identified over time. We might be completely anonymous in a system that uses unique cookies to track us as we browse the internet, but the same process of correlation and discrimination still occurs. It’s the same with faces; we can be tracked as we move around a store or shopping mall, even if that tracking isn’t tied to a specific name.”[126]

The Office of the Privacy Commissioner of Canada (OPCC) has also taken a broad, contextual view of the definition of personal information, when considering online behavioural advertising (OBA). Although noting that this view about identifiability should be tested on a case-by-case basis, the OPCC’s view is that:

> “in the context of OBA, given the fact that the purpose behind collecting information is to create profiles of individuals that in turn permit the serving of targeted ads; given the powerful means available for gathering and analyzing disparate bits of data and the serious possibility of identifying affected individuals; and given the potentially highly personalized nature of the resulting advertising, it is reasonable to take the view that the information at issue in behavioural advertising not only implicates privacy but also should generally be considered ‘identifiable’ in the circumstances”.[127]

The OPCC has further noted that their consultations found general agreement amongst stakeholders that “there were privacy implications from online tracking (even if not all agreed that the data collected from tracking was personal information)”.[128] Whether or not identity could be determined from the information gleaned through online tracking was the sticking point, in terms of meeting the legal definition for ‘personal information’.

If the end result of an activity is that an individual can be *individuated* from a dataset, and targeted for some intervention, that is a privacy harm that may need protecting against. (Whether or not it is actually deserving of protection will depend on the context; for example an intervention for fraud prevention or crime detection is a different proposition to online behavioural advertising. It is within the APPs that the law defines the allowable purposes for the collection, use or disclosure of personal information, such as for fraud prevention).

[126] Bruce Schneier, “We’re banning facial recognition. We’re missing the point”, *New York Times*, 20 January 2010; available at https://www.nytimes.com/2020/01/20/opinion/facial-recognition-ban-privacy.html

[127] Office of the Privacy Commissioner of Canada, “Policy position on online behavioural advertising”, December 2015; available at https://www.priv.gc.ca/en/privacy-topics/technology/online-privacy-tracking-cookies/tracking-and-ads/bg_ba_1206/

[128] Office of the Privacy Commissioner of Canada, “Policy position on online behavioural advertising”, December 2015; available at https://www.priv.gc.ca/en/privacy-topics/technology/online-privacy-tracking-cookies/tracking-and-ads/bg_ba_1206/

In a digital world, ‘not identified’ is no longer an effective proxy for ‘will suffer no privacy harm’. Individuation must be anticipated by privacy laws as well.

A new definition

While the concept of individuation might appear novel at first, it is simply an evolution of the concept of identifiability.

By appreciating that the concept of ‘identifiability’ is not limited to the ability to know or verify a person’s legal identity, but that it can also encompass the act of ‘singling out’ – i.e. disambiguating one individual from another by way of either recognising or assigning certain characteristics to each person - we are already embracing the concept of individuation.

The concept of individuation is arguably already included in the GDPR, given:

- the definition of ‘personal data’ includes the concept of both direct and ‘indirect’ identification[129]

- online identifiers and location data are explicitly mentioned within the definition of ‘personal data’ as possible identifiers, by reference to which an individual might be “identified, directly or indirectly”.

- the concept of online identifiers is explicated further in Recital 30 to include IP address, cookies and RFID tags where used to create profiles of people and identify them;[130] and

- Recital 26 mentions ‘singling out’ as a means by which someone might become ‘identifiable’.[131]

Professors Paterson & McDonagh, referring to Recital 26 in the GDPR, conclude:

> “The express reference to ‘singling out’ suggests that the processing of data that singles out but does not reveal an individual’s identity comes within the scope of European data protection law”.[132]

[129] Article 4, *General Data Protection Regulation*, Regulation 2016/679 of the European Parliament and of the Council

[130] Recital 30 of the GDPR states: “Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them.”

[131] Recital 26 of the GDPR includes: “To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly.”

[132] Moira Paterson & Maeve McDonagh, “Data Protection in an era of Big data: the Challenges posed by Big Personal Data”, *Monash University Law Review*, Vol 44(1), 2018, p.16; available at https://www.monash.edu/__data/assets/pdf_file/0009/1593630/Paterson-and-McDonagh.pdf ; see also Dr Frederik J. Zuiderveen Borgesius, “Singling out people without knowing their names – Behavioural targeting, pseudonymous data, and the new Data Protection Regulation”, Computer Law & Security Review, Vol 32(2), April 2016, pp.256-271; available at https://www.sciencedirect.com/science/article/pii/S0267364915001788?via%3Dihub

Even newer statutes and other international instruments have shifted towards incorporating the concept of individuation, moving beyond just identifiability as the essential threshold element:

- The 2018 Californian privacy law CCPA expressly includes, within its definition of personal information, data which is "**capable of being associated with**, or could reasonably be **linked**, directly or indirectly, with a particular consumer or household"; and includes within its definition of 'unique identifier': "a persistent identifier that can be used **to recognize** a consumer, a family, or a device that is linked to a consumer or family, over time and across different services, including, but not limited to, a device identifier" (emphasis added).

- The 2019 Nigerian privacy statute explicitly defines 'identifiable information' to include "information that can be used on its own or with other information to **identify, contact, or locate** a single person, or to identify an individual in a context" within its definition of 'personal data' (emphasis added).

- The 2019 international standard in Privacy Information Management, ISO 27701, incorporates data which could be "**directly or indirectly linked**" to an individual, *regardless* of whether the individual can be identified, within its definition of 'personally identifiable information' (emphasis added).

So as to enable clarity and consistency in the application of privacy law, and to protect against the potential privacy harms enabled by individuation, this Research Paper concludes that the Privacy Act should incorporate a definition for the word 'identifiable', as follows:

> "(i) able to be identified, *or* (ii) able to be discerned or recognised as an individual distinct from others, regardless of whether their identity can be ascertained or verified"

We further suggest that the definition should clearly provide that the test for identifiability is that an individual will be considered "able to be discerned or recognised as an individual distinct from others":

> "if the individual, or a device linked to the individual, could (whether online or offline) be surveilled, tracked or monitored; or located, contacted or targeted; or profiled in order to be subjected to any action, decision or intervention including the provision or withholding of information, content, advertisements or offers; or linked to other data which is about or relates to the individual"

This additional layer to the test for identifiability aims to ensure that the scope of the regulation does not over-reach into technologies which do *not* pose risks of privacy harms, such as the use of sessional or load-balancing cookies which are necessary to make a website work, but which do not then continue to track the user.

The impact of this reform would include that the act of placing a tracking cookie on a person's connected device, and then tracking that person's online behaviour in a way which distinguished them from other individuals, in order to profile and then target that person (for example, in order to serve up an advertisement, or to determine what offers or pricing to show that person) will constitute the handling of personal information, such that the privacy principles apply to that conduct, notwithstanding that the advertiser and online ad broker could each claim not to know 'the real identity' of the person.

By more explicitly embedding the concept of individuation within the core definitional element of 'identifiability', the Privacy Act will be evolving in the same direction as other recent privacy laws such as the GDPR and CCPA.

Resolving pragmatic issues

A critical consideration for any proposed reform to the definition of personal information is how it would work in practice. In the context of this law reform proposal, a further consideration is how any reform would work in light of the other reforms recommended by the ACCC, such as increasing the coverage of the Privacy Act to include small businesses.

As noted above, the proposed test for identifiability includes that an individual will be considered "able to be discerned or recognised as an individual distinct from others" only in certain circumstances. The intent of the definition is to capture activities which could lead to privacy harms, and exclude activities which could not.

An example of a business activity which would not be included under the proposed new definition is that of a small business website which collects the IP addresses of users for basic analytics purposes (e.g. to see the countries where users are purportedly located) but which has no capability to track a user from their IP address. This balances the need for privacy to be protected regardless of where a potential privacy harm might arise, while not over-burdening smaller organisations whose practices pose low to no risk.

A further pragmatic issue to be resolved is how to ensure that the revised definition of personal information will work in practice with the privacy principles.

APPs 12-13 are the Access and Correction principles. If an individual has not been 'identified' by an organisation, but has only been 'singled out', how could the individual verify themselves such as to exercise their access or correction rights against that organisation?

This practical dilemma is not entirely novel. The definition of personal information includes information "whether the information or opinion is recorded in a material form or not". If information is not recorded in a material form – for example, it has only been observed by an employee of an agency – then how can the individual seek access or correction? The answer is they can't. The Access and Correction principles already only work in practice in relation to a sub-set of personal information, namely personal information that *has* been recorded in a material form. (By contrast, other privacy principles still make sense with respect to unrecorded information; gossiping about what a client did or looks like, without anything ever being recorded, could still constitute an unauthorised disclosure in breach of APP 6.)

This Research Paper suggests that an expanded scope for definition of personal information should not be rejected just because access and correction rights cannot be realised for *some* types of personal information – a pre-existing problem in any case. The privacy of the individuals affected by individuation is still worthy of protection.

Whether in relation to access or correction, or any additional GDPR-type data subject rights to be introduced as part of the wider reform of the Privacy Act (such as a right to erasure, a right to algorithmic transparency, etc), the management of data subject rights needs careful consideration in terms of the degree of identity verification needed from an applicant. Regulated entities should avoid a situation in which APP 12 could be weaponised by a perpetrator of family violence to impersonate their ex-partner and seek access to geolocation data; or by any other motivated intruder seeking to find out personal information about their target. This is an existing issue in relation to access and correction requests.

On the other hand, the drafting of the definition of personal information should not be so prescriptive about being able to identify or single out *individuals* that it provides entities an excuse not to comply with their privacy obligations at all. For example, the privacy legal framework should not be frustrated by an electricity provider claiming smart energy meter data is not personal information (and thus they don't need to protect it at all) just because they cannot single out an *individual's* use of electricity from the rest of their *household*. This consideration affects all the APPs. (There is further discussion about the privacy of *households* below.)

This Research Paper offers the following proposed amendments by way of solution:

- APPs 12-13 (and any other new data subject rights) will only apply to *recorded* personal information, where the applicant is able to demonstrate to an appropriate degree of certainty (commensurate with the level of risk posed by an unauthorised disclosure of the data sought) that they are either (i) the individual to whom the data relates or is about, or (ii) a nominated account holder on behalf of a household account, or (iii) authorised to represent all members of a household

- APPs 2-11 will apply notwithstanding that an entity cannot discern or recognise an individual as distinct from other members of the same household

- Provisions relating to making and responding to a complaint, the OAIC's powers, the data breach notification scheme and all other provisions relating to remedies, will apply notwithstanding that an individual complainant cannot demonstrate that the data at issue relates solely to themselves, as distinct from other members of the same household

Devices and households

This section examines whether devices or households should be specifically included within the definition of personal information.

Devices

With the exception of the CCPA, none of the privacy laws reviewed for this Research Paper cover information about or drawn from devices as personal information *per se*. However many jurisdictions, whether through explicit mention in the statute, in case law or regulatory guidance, recognise that to the extent that a *device* can be identified (this unique mobile phone, that unique smart TV):

- a device identifier can be a proxy for an identifier for the individual who is the user of that device, and
- information about the use or physical movements of the device is information 'about' or 'relating to' the behaviour or physical movements of the individual who is the user of that device

and therefore to the extent that an individual is 'identifiable' via a device, the definition of 'personal information' (or its equivalent) can be met.

The CCPA is similar, with devices being mentioned in the context of unique identifiers, probabilistic identifiers, and aggregate consumer information (a test for de-identification). The effect of the whole is that the CCPA makes clear that devices are a means by which data may be collected from, or attributed to, particular consumers (or households, or families).

Devices are also referenced in the proposed CCPA Regulations as a means by which privacy messaging can be communicated to consumers.

The key difference between CCPA and other privacy laws in terms of its treatment of devices is that one of the criteria for businesses to be regulated by the CCPA in the first place is if a business:

> "annually buys, receives for the business's commercial purposes, sells, or shares for commercial purposes, alone or in combination, **the personal information of 50,000 or more** consumers, households, or **devices**" (emphasis added).[133]

[133] CCPA section 1798.140(c)(1)(B)

The phrase “or devices” in that section is either redundant (because personal information rights are only actionable by consumers or households), or the intent of the section was to draw into the scope of regulation any business which sells more than 50,000 devices, *regardless* of whether personal information is also collected. Given that ‘device’ is defined to mean any physical object capable of connecting to another physical object,[134] the latter interpretation could in theory capture a business which did nothing more than sell 50,000 pieces of Lego for cash to an anonymous buyer. We assume this was not the intention of the drafters.

This Research Paper takes the view that the phrase “or devices” in section 1798.140(c)(1)(B) of the CCPA is likely a drafting error, and is redundant.

Under Australian law, the effect of s.187LA of the TIA Act, described above, is to encompass a protective notion about the privacy of communications, which recognises that information about devices can be a proxy for an individual (or, indeed, for households), who or which should be afforded some privacy rights and privacy protections, as a balance against the more intrusive aspects of the metadata retention scheme. It has the effect of bringing all ‘retained data’ under the umbrella of ‘personal information’, as a way of delivering those privacy rights and protections.

The Australian Parliament has therefore already recognised that a device can be a proxy for the individual or household using that device, and that therefore data about (or collected from) devices is deserving of privacy protection.

This Research Paper takes the view that there is no need for ‘devices’ to be recognised as a *category* of personal information. Instead, the recommendations below suggest that devices be called out, but only in the context of establishing the test for identifiability of an individual.

We also suggest including a drafting note to the effect that “device” is to be read expansively, and can include a vehicle such as a car, a mobile device such as a mobile phone, a wearable such as a FitBit, an implantable such as a pacemaker, or a household device such as a smart TV. We further suggest that, unlike the CCPA, the definition should *not* limit the definition with reference to the connectivity of devices. A vehicle could be tracked via aerial surveillance, or a photograph of its number plate, without needing the vehicle to ‘connect’ to the internet or to any other device.

Households

With the exception of the CCPA, it appears that privacy laws which specifically mention households or families do so not so much to give legal recognition to the rights of any particular type of grouping of individuals, but simply as a way of overcoming what would otherwise be a barrier to the exercise of privacy rights or the provision of privacy protections,

[134] CCPA section 1798.140(j) defines device as “any physical object that is capable of connecting to the internet, directly or indirectly, or to another device”,

where data can be linked to a dwelling, but does not enable further ‘identification’ of an individual within that dwelling. Examples include where data is collected from household devices such as smart electricity meters, or connected consumer devices, the use of which is shared or blurred between different members of a household.

In European jurisprudence, for example, the notion of ‘indirect’ identification of an individual has been interpreted to enable privacy rights for an individual where the holder of the data could only identify which property the data related to.[135]

Likewise in New Zealand, the Privacy Commissioner concluded that information about a home Wi-Fi network is personal information if it is possible to tell that a network is located in a particular person’s home;[136] and that electricity usage measured by a smart meter is personal information once it is linked to an account holder, even if the precise household member using the electricity could not be identified.[137] This view was confirmed by the New Zealand Supreme Court: power consumption data was ‘personal information’ “in the sense that it indicates the power consumption at a place owned and occupied by identifiable individuals”.[138]

The CCPA has taken this notion a step further, by including in its definition of personal information: “information that identifies, relates to, describes, is capable of being associated with, or could reasonably be linked, directly or indirectly, with **a particular consumer or household**” (emphasis added).[139]

However, the devil is in the detail: pragmatic but complex rules as to how to define a ‘household’, and how to enable privacy rights such as the right of access for a household, were left to the regulations. Despite the CCPA commencing on 1 January 2020, those regulations, as at the date of writing, are still only in the form of an exposure draft, on which the Californian Attorney General is currently seeking submissions.

The draft regulations provide:

> “‘Household’ means a person or group of people who: (1) reside at the same address, (2) share a common device or the same service provided by a business, and (3) are identified by the business as sharing the same group account or unique identifier.”[140]

Confusingly, the CCPA separately refers to unique identifiers which are linked to, or can be used to recognise, “a consumer or **family**” (rather than *household*); family is then defined to mean minor children and their parent/guardian. Neither the Act nor the proposed Regulations make any attempt to reconcile these two competing groupings of individuals.

[135] Lee Bygrave, *Data protection law: approaching its rationale, logic and limits*, Kluwer Law International, 2002, part 2.4.1.
[136] See the 2010 inquiry into Google StreetView, available at https://www.privacy.org.nz/news-and-publications/commissioner-inquiries/google-s-collection-of-wifi-information-during-street-view-filming/
[137] Case Note 251185 [2015] NZ PrivCmr 3; available at https://www.privacy.org.nz/news-and-publications/case-notes-and-court-decisions/case-note-251185-2015-nz-privcmr-3-use-of-smart-meters-by-utility-companies/
[138] R v Alsford [2017] NZSC 42 at [30]; available at https://www.courtsofnz.govt.nz/assets/cases/2017/d2lj.pdf
[139] CCPA section 1798.140(o)(1)
[140] Text of Modified Regulations [Clean Version] Title 11. Law Division 1. Attorney General Chapter 20. California Consumer Privacy Act Regulations Proposed Text Of Regulations, part 999.301(k); available at https://www.oag.ca.gov/sites/all/files/agweb/pdfs/privacy/ccpa-text-of-mod-clean-020720.pdf

By including households (or, in some contexts, families) within the definition of personal information, the CCPA has the effect of extending all privacy rights and protections available under that Act to households (or families) as well as individuals. This is causing some difficulties in practice, both for regulated businesses and for the Attorney General of California.

For example, how can a 'household' exercise its rights of access/correction (or other rights under the CCPA such as deletion, or opting out, aka 'do not sell')? How is a household to verify its *bona fides* to a business which is expected to provide the household with access to (or deletion of) the 'personal information' held about the household? What if one family member uses their access to 'household' data to harm another family member? If notice is to be provided about a collection of personal information, how should the notice address a household?

However this is not a novel complication in privacy law. Caselaw[141] and guidance from the OAIC provide various examples of information which is so intertwined that it may constitute the personal information of more than one person, such as a statement that Sue was born with foetal alcohol syndrome which reveals information about Sue *and* about Sue's mother.[142]

The proposed Regulations to be made under the CCPA to manage access and deletion rights are, as at the date of writing, that:

> "(a) Where a household does not have a password-protected account with a business, a business shall not comply with a request to know specific pieces of personal information about the household or a request to delete household personal information unless all of the following conditions are satisfied:
>
> > (1) All consumers of the household jointly request access to specific pieces of information for the household or the deletion of household personal information;
> >
> > (2) The business individually verifies all the members of the household subject to the verification requirements set forth in section 999.325; and
> >
> > (3) The business verifies that each member making the request is currently a member of the household.
>
> (b) Where a consumer has a password-protected account with a business that collects personal information about a household, the business may process requests

[141] See *Jackson v The University of New South Wales* [2018] NSWCATAD 12 at [90]; *AFC v The Sydney Children's Hospital Specialty Network (Randwick and Westmead)* [2012] NSWADT 189 at [38-39]; and *AJD v Royal Prince Alfred Hospital* [2014] NSWCATAD 125 at [73].

[142] Office of the Australian Information Commissioner, *What is personal information?*, May 2017; available at https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information

> to know and requests to delete relating to household information through the business's existing business practices and in compliance with these regulations.
>
> (c) If a member of a household is a minor under the age of 13, a business must obtain verifiable parental consent before complying with a request to access specific pieces of information for the household or the deletion of household personal information pursuant to the parental consent provisions in section 999.330".[143]

These proposed rules do however lead to further questions, such as how a business might be expected to know how many people live at a residence (let alone how to verify each person's identity), or whether there is anyone under the age of 13 living at a residence?

Consumers have complained that in order to assert their privacy rights under the CCPA, they are actually being forced to hand over *more* personal information about themselves to companies than those companies already had.[144] On the other hand, if companies do not verify identities accurately, they risk privacy harm by making unauthorised disclosures to the wrong person, as occurred in some cases soon after the GDPR's expanded data subject rights scheme went into effect.[145]

This Research Paper offers the position that there is no need to expand the scope of the Privacy Act to include 'households' or 'families' as the holders of privacy rights, but that the concept of a household is relevant to identifiability, and how the privacy principles and other provisions of the Privacy Act should operate in practice when data can be distinguished down to the household level but no further. This topic is addressed above, under *Individuation*.

[143] Text of Modified Regulations [Clean Version] Title 11. Law Division 1. Attorney General Chapter 20. California Consumer Privacy Act Regulations Proposed Text Of Regulations, part 999.318; available at https://www.oag.ca.gov/sites/all/files/agweb/pdfs/privacy/ccpa-text-of-mod-clean-020720.pdf

[144] Alistair Barr, "Come on a trip into the new privacy circle of hell", *Bloomberg*, 9 January 2020; available at https://www.bloomberg.com/news/newsletters/2020-01-09/come-on-a-trip-into-the-new-privacy-circle-of-hell

[145] Kashmir Hill, "Want Your Personal Data? Hand Over More Please", *New York Times*, 15 January 2020; available at https://www.nytimes.com/2020/01/15/technology/data-privacy-law-access.html

Should examples be enumerated in statute?

One of the matters to be resolved when considering a re-draft of the definition of personal information is whether the statute should include a list of examples of what constitutes personal information. Mentioning 'technical data', metadata or location data within a list of examples of what constitutes personal information is one of the possible ways by which to give force to the ACCC's recommendation.

Laws which feature a list of examples include:

- Tasmania
- Canada (although only the Act which regulates the public sector)
- South Africa, and
- the CCPA.

However presenting a list of items in statute moves away from the contextual analysis approach favoured to date in Australian privacy statute law and case law (with the exception of the Tasmanian statute). This can undermine the purpose of the elements of the definition, by suggesting that the enumerated examples are *de facto* 'about' an 'identifiable' individual. In reality such data might *not* be identifiable if it is presented in a certain way, such as if it has been aggregated, subject to techniques such as differential privacy, or is protected by techniques such as homomorphic encryption or salted hashing.

Further, presenting a list of examples encourages readers not to think beyond the items on the list, to robustly apply the elements of the definition itself. A prescriptive list can draw readers' attention away from the fuller intention of the definition, in which case nuances can be lost.

In its detailed consideration of the Privacy Act in 2008, the Australian Law Reform Commission – and the Australian Government in acting on its recommendations – opted to maintain a flexible and technology-neutral approach to defining the scope of statutory privacy protections. Any enumeration of examples risks negating that technology-neutral approach. As noted above, HIPAA offers a useful lesson in the risks posed by being too prescriptive or calling out particular technologies within the wording of statute law.

While this Research Paper presents a number of recommendations for the inclusion of drafting notes within the statute itself, it does not recommend enumerating examples within the definition of personal information.

Conclusions

Meeting the demands of the digital economy

Data is the lifeblood of the digital economy, and will increasingly power decision-making in all sectors of the economy.

Robust data protection regulation is necessary to achieve both consumer protection outcomes, and consistency of the playing field for industry. It will therefore be critical to ensure that the Privacy Act remains fit for its purpose of enabling effective regulation of personal information handling, in line with community and business expectations.

The ACCC already identified that the current definition suffers from a lack of certainty around its coverage of technical data. The OAIC also raised the issue of whether inferred data is within scope, while a multiplicity of stakeholders responded to the Digital Platforms Inquiry to raise their concerns that privacy regulations were not keeping up with the challenges posed by new technologies.

This Research Paper has found that in order to offer protection from privacy harms, the law needs to recognise that in a digital world, ‘not identified’ is no longer an effective proxy for ‘will suffer no privacy harm’. Privacy laws must anticipate the harms that can arise via individuation, or ‘singling out’, as well.

By more explicitly embedding the concept of individuation within the core definitional element of ‘identifiability’, the Privacy Act can be modernised to reflect the reality of the digital economy.

Global consistency

As is evidenced in this Research Paper, there is considerable variation in the scope afforded by different jurisdictions’ definitions of ‘personal information’ (or ‘personal data’, etc).

While the GDPR might be considered to currently set the global benchmark in terms of the strength of privacy law, in reality with respect to *some* aspects of the definition, other laws go further than the GDPR. Examples include laws which explicitly cover households (e.g. CCPA), and the privacy of deceased persons (e.g. most of the State and Territory privacy laws in Australia).

Nonetheless, there is considerable benefit in seeking to align the Privacy Act to the GDPR in particular.

First, we understand anecdotally, based on the author's discussions with clients including businesses from small to large, as well as with privacy regulators, that the expectation from Australian industry is that consistency with the GDPR would be preferable. Given GDPR's extra-territorial reach, many organisations in Australia have had to expend considerable compliance effort in relation to GDPR. Greater consistency between the Australian Privacy Act and the GDPR would assist organisations to implement a single, consistent global compliance program, instead of having to 'localise' policies and procedures, internal and external-facing messaging, and staff training.

Second, given the degree of media focus on the GDPR and newer rights such as erasure and algorithmic transparency, closer alignment with the GDPR will assist consumers and citizens to understand and assert their privacy rights.

Third, closer alignment to the GDPR will assist the OAIC in joint investigations with other regulators, where conduct affects many consumers or citizens across jurisdictions. It will also assist privacy regulators to promote the adoption of privacy management frameworks which can operate interchangeably between different jurisdictions.

Fourth, there could be an economic benefit for Australia in seeking greater consistency with the GDPR. The trade-off for strengthening the robustness of privacy regulation on businesses regulated under the Australian Privacy Act is the economic benefit that would accrue by creating the right conditions for an EU 'adequacy' decision. Unlike some of our Asia-Pacific neighbours such as New Zealand, Japan, Canada and (to a lesser extent) the United States,[146] Australia is not currently recognised by the European Commission as having a privacy legal framework equivalent to that of the EU.

Without an 'adequacy' decision, Chapter V of the GDPR has the effect of prohibiting the movement of personal data out of EU countries to Australia, unless certain stringent conditions are met. In practice, meeting those conditions is extremely time-consuming and costly for Australian businesses. By contrast, obtaining an 'adequacy' decision could drive innovation and growth of the digital economy, especially for Australian small to medium-sized businesses seeking to access EU markets.

[146] In the US only companies which have current Privacy Shield certification are recognised as 'adequate'. The Canadian adequacy decision is limited to private sector organisations only, not public sector agencies. South Korea was being assessed as at the date of writing. The list of adequacy decisions is at https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en#dataprotectionincountriesoutsidetheeu

Proposed reforms

Our recommendations overleaf have therefore been drafted with the following objectives:

(a) Resolving the lack of clarity around coverage of technical data, in line with the ACCC's recommendations

(b) Incorporating elements of the definition which are now relatively uncontested, but which for clarity's sake should be made plain in the statute

(c) Enabling consistency with more recent developments in Australian privacy jurisprudence such as the CDR scheme and the telecommunications data retention scheme

(d) Enabling consistency with the GDPR where suitable, and

(e) Updating the definition to ensure it remains fit for purpose in digital environments.

Recommendations

1. **Amend the definition of 'personal information' in the Privacy Act as follows:**

    a. change "about" to "which is about or relates to"

> **Rationale**
>
> This amendment would partially help to resolve the lack of clarity around coverage of technical data, in line with the ACCC's recommendations.
>
> It would also bring this definitional element closer into alignment with the CDR scheme and the telecommunications data retention scheme, as well as the GDPR and other international privacy laws.

b. include a drafting note or additional definition to further explicate that "about or relates to" means

(i) if the individual is a subject of the information, or

(ii) if the information concerns or links to the individual, or

(iii) if the intent or effect of the information's handling will be to learn, evaluate, treat in a certain way, make a decision about, influence the status or behaviour of, or otherwise have an impact upon, the individual.

> **Rationale**
>
> This proposal is to avoid judicial determinations which might otherwise interpret the scope of the phrase with the same narrow 'subject matter' focus much criticised in the *Telstra* case.

c. make the phrase "identified or identifiable individual"

> **Rationale**
>
> This would replace the current wording of "identified individual, or an individual who is reasonably identifiable".
>
> Removing the concept of *'reasonably'* identifiable brings the definition into line with the GDPR and indeed every other national privacy law and international instrument surveyed.

d. add the phrase “whether the information or opinion is provided, collected, created, generated or inferred”

Rationale

This amendment will make explicit that inferences drawn from data, or new data generated about an individual (such as customer insights which lead to ratings being applied to a customer profile), are within the scope of the definition.

Rationale

Following the above recommendations will result in a revised definition as follows:

> ***"personal information"*** means information or an opinion which is about or relates to an identified or identifiable individual:
>
> (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; and
>
> (c) whether the information or opinion is provided, collected, created, generated or inferred

2. **Introduce a definition for ‘identifiable’, which includes the following elements:**

Rationale

By defining ‘identifiable’ separately to ‘personal information’, a number of problems with interpreting the current definition of ‘personal information’ can be resolved, without overly complicating the definition of ‘personal information’ itself.

A number of international privacy statues, including the GDPR, define or explain the word ‘identifiable’.

a. the core element of identifiability to be defined as “(i) able to be identified, *or* (ii) able to be discerned or recognised as an individual distinct from others, regardless of whether their identity can be ascertained or verified”

Rationale

By embedding the concept of individuation within the phrase ‘identifiable’, the overarching legal framework is not altered, while achieving the desired level of protection against the types of privacy harms made more prevalent in a digital environment.

b. provide that the test for identifiability is that an individual will be considered 'able to be discerned or recognised as an individual distinct from others': "if the individual, or a device linked to the individual, could (whether online or offline) be surveilled, tracked or monitored; or located, contacted or targeted; or profiled in order to be subjected to any action, decision or intervention including the provision or withholding of information, content, advertisements or offers; or linked to other data which is about or relates to the individual"

Rationale

This proposal aims to further explicate the notion of identifiability by individuation, by going to the heart of the types of privacy harms which can occur even when the precise 'identity' of an individual is unknown by the perpetrator of the harm. Mentioning decisions in relation to the provision or withholding of information is intended to capture the curation and delivery of 'personalised content' such as ads, price offers, news feeds, recommendations for related content, etc. The definition should cover decisions to *exclude* people from seeing certain content as much as it covers decisions to target or include people: e.g. a decision *not* to show a particular job ad to people outside a certain age bracket, or who identify as (or who have been inferred as belonging to) a particular ethnicity or religion.

The proposed language also aims to ensure that the scope of the regulation does not over-reach into technologies which do *not* pose risks of privacy harms, such as the use of sessional or load-balancing cookies which are necessary to make a website work, but which do not then continue to track the user; or a small business website which collects the IP addresses of users for basic analytics purposes (e.g. to see the countries where users are purportedly located) but which has no capability to track a user from their IP address.

c. provide that the test for identifiability is to be conducted with consideration to the information either alone or in combination with other available information

Rationale

This incorporates the 'linkability' element of the definition which is now relatively uncontested, but which for clarity's sake should be made plain in the statute.

d. provide that the test for identifiability is not to be conducted with consideration only to whether the individual is identifiable to the entity collecting or holding the personal information, but must anticipate the likelihood of identifiability being achieved by any intended or likely recipients of the data

Rationale

This incorporates an element of the definition which is now relatively uncontested, but which for clarity's sake should be made plain in the statute.

e. include a drafting note to the effect that location data, device identifiers and online identifiers (including cookies, IP addresses, MAC addresses, user IDs), are examples (but not the only examples) of data, identifiers or techniques which can render an individual able to be discerned or recognised as an individual distinct from others

Rationale

This amendment would further help to resolve the lack of clarity around coverage of technical data, in line with the ACCC's recommendations; and would also help to achieve consistency with the GDPR and privacy laws in other jurisdictions featuring more modern statues (such as South Africa and Nigeria).

The Explanatory Memorandum for the amending Bill could offer further examples of how an individual might be considered 'identifiable' even if their identity is not known, including offline examples.

f. include a drafting note (or add a new definition to the Act) to the effect that "device" is to be read expansively, and can include a vehicle such as a car, a mobile device such as a mobile phone, a wearable such as a fitness tracker or location monitor, an implantable such as a pacemaker, or a household device such as a smart TV.

Rationale

This amendment would further help to resolve the lack of clarity around coverage of technical data, in line with the ACCC's recommendations; and would also help to achieve protection against the types of privacy harms made more prevalent in a digital environment.

The definition should *not* limit the definition with reference to the connectivity of devices.

The scope of 'personal information' will still be limited by the requirement that the device is somehow linked to an individual, and that the individual is somehow identifiable.

3. **Replace the definition of ‘de-identified’ in the Privacy Act**:

    a. add a new definition: “anonymous data means data from which no individual is identifiable”

Rationale

This wording offers simplicity and clarity, and consistency with the intent of the GDPR. It also makes clear that to fall within the bounds of this definition (and thus have the benefit of no longer needing to protect the data under privacy laws), *no* individual can be identifiable from the data. This protects both the passenger and the driver in the taxi trip dataset scenario.

The separate definition of ‘identifiable’ makes it clear that the test is to be considered by looking at the information in the context of possible linkages with other data sources, rather than in isolation.

We do *not* recommend word-for-word adoption of the GDPR definition, because it contains a notable weakness: by focussing on the ‘data subject’, data meeting the GDPR definition could still potentially reveal information about a third party (e.g. the passenger of a taxi in the taxi trip dataset scenario, because only the taxi *driver* would be the ‘data subject’).

Nonetheless using the term ‘anonymous data’, instead of ‘de-identified’ would align the Privacy Act with the GDPR.

This amendment would also achieve greater clarity for regulated entities. A description of how data has been treated in a statistical sense should not be conflated with an assertion about its legal status under the Privacy Act. By using a word which is an adjective (‘anonymous’), instead of a word which can be either an adjective or a verb (‘de-identified’), this problem of conflation between the two meanings should be ameliorated, if not completely resolved.

Indeed we counsel against using the word ‘de-identified’ at all, instead preferring the descriptor “data to which de-identification techniques have been applied”.

    b. include a drafting note to the effect that relevant parts of the Privacy Act (namely, the APPs and the data breach notification scheme) therefore do not apply to the handling of anonymous data.

Rationale

This amendment would achieve greater clarity for regulated entities and consumers/citizens alike, and would align the Privacy Act with the GDPR. The drafting note should be similar to that found in Recital 26 of the GDPR.

Either the drafting note or the Explanatory Memorandum for the amending Bill could offer further explication, stating that just because de-identification techniques have been applied to data will not necessarily make the data ‘anonymous’ in the legal sense required here. In particular, it should be clarified that pseudonymous data is *not* anonymous data; see Article 4 and Recital 26 of the GDPR.

4. **Introduce amendments to clarify the practical operation of various elements of the Privacy Act:**

   a. APPs 12-13 (and any other new data subject rights) will only apply to *recorded* personal information, where the applicant is able to demonstrate to an appropriate degree of certainty (commensurate with the level of risk posed by an unauthorised disclosure of the data sought) that they are either (i) the individual to whom the data relates or is about, or (ii) a nominated account holder on behalf of a household account, or (iii) authorised to represent all members of a household

   b. APPs 2-11 will apply notwithstanding that an entity cannot discern or recognise an individual as distinct from other members of the same household

   c. Provisions relating to making and responding to a complaint, the OAIC’s powers, the data breach notification scheme and all other provisions relating to remedies, will apply notwithstanding that an individual complainant cannot demonstrate that the data at issue relates solely to themselves, as distinct from other members of the same household

Rationale

These provisions aim to alleviate the burden upon regulated entities – particularly small businesses – facing access or correction requests from individuals with whom they do not already have a customer relationship through which verification of the applicant’s *bona fides* may be established.

These provisions also aim to ensure that the drafting of the definition of personal information is not so prescriptive about being able to identify or single out *individuals* that it provides entities an excuse not to comply with their privacy obligations at all, where their relationship is with a *household* of individuals.

5. **The scope of 'personal information' should be expanded to cover deceased persons for a period of 30 years after death**.

Rationale

Subject to further consideration and consultation on this point, we recommend that the protections and obligations afforded under the Privacy Act should be extended to deceased persons for a period of 30 years after death.

This proposal would take the Privacy Act beyond equivalence with the GDPR, which only covers living persons. However coverage of the deceased is already a feature of privacy law in most States and the two Territories in Australia. 30 years would be a period consistent with NSW and Victorian privacy laws, as well as Cabinet confidentiality periods under the *Archives Act 1983* (Cth).

6. **Advice should be sought from Parliamentary Counsel as to whether any cognate amendments should be made to the TIA Act, and/or whether the 'Note' referring to the TIA Act should be removed from the definition of personal information, once the above recommended amendments have been drafted**.

Case study: Metadata, identifiability and the meaning of 'about'

As a technology journalist, Ben Grubb was interested to find out what the geolocation data collected by his mobile phone service provider could reveal about him. He wanted to replicate the efforts of a German politician,[147] to illustrate the power of geolocation data to reveal insights into not only our movements, but our behaviour, intimate relationships, health concerns or political interests.[148]

Exercising his rights under what was then National Privacy Principal (NPP) 6.1, Grubb sought access from his mobile phone service provider, Telstra, for his personal information – namely, "all the metadata information Telstra has stored about my mobile phone service (04…)".

Telstra refused access to various sets of information, including location data on the basis that it was not 'personal information'. Grubb lodged a complaint with the Australian Privacy Commissioner. While the complaint was ongoing, Telstra provided a folder of billing information, outgoing call records, and the cell tower location information for his mobile phone at the time when he had originated a call, which is data kept in its billing systems.

What was not provided, and what Telstra continued to argue was not 'personal information' and thus need not be provided, included 'network data'. Telstra argued that that geolocation data – the longitude and latitude of mobile phone towers connected to the customer's phone at any given time, whether the customer is making a call or not – was not 'personal information' about a customer, because on its face the data was anonymous.

In Grubb v Telstra,[149] the Privacy Commissioner found that journalist Ben Grubb was entitled to access the 'metadata' held about him by his mobile phone service provider.

The Privacy Commissioner found that a customer's identity *could* be linked back to the geolocation data by a process of cross-matching different datasets. (Indeed, the decision notes that Telstra conducted such cross-matching approximately 85,000 times in 2013-14, in order to provide data to law enforcement agencies.)

Telstra appealed against the Privacy Commissioner's determination, and the matter went to the Administrative Appeals Tribunal (AAT).

[147] Kai Biermann, "Betrayed by our own data", *Zeit Online*, 10 March 2011; available at https://www.zeit.de/digital/datenschutz/2011-03/data-protection-malte-spitz
[148] Ben Grubb, "Spies can access my metadata, so why can't I? My 15-month legal battle with Telstra", *Sydney Morning Herald*, 10 October 2014; available at https://www.smh.com.au/technology/spies-can-access-my-metadata-so-why-cant-i-my-15month-legal-battle-with-telstra-20141010-1146go.html
[149] *Ben Grubb and Telstra Corporation Limited* [2015] AICmr 35 (1 May 2015); available at http://www8.austlii.edu.au/cgi-bin/viewdoc/au/cases/cth/AICmr/2015/35.html

Once in the AAT, the nature of the case shifted. Both parties' submissions argued about how the definition of 'personal information' should be interpreted, in terms of whether or not Grubb was 'identifiable' from the network data, including how much cross-matching with other systems or data could be expected to be encompassed within the term 'can reasonably be ascertained' (which was the test contained the definition of 'personal information' at the time). However the AAT drew no solid conclusion about whether or not Grubb was actually identifiable from the network data in question.

Instead, the AAT questioned whether the information was even 'about' Grubb at all.

The AAT found that there was a two-step process to meeting the definition of personal information; the information must be *about* an individual, and in a separate inquiry, that individual must be reasonably identifiable from that information. The AAT stated:

> "The starting point must be whether the information or opinion is about an individual. If it is not, that is an end of the matter and it does not matter whether that information or opinion could be married with other information to identify a particular individual".[150]

In other words, the AAT's position was that the fact the information might relate or link back to an individual does not necessarily make it "about" that individual. The AAT quoted a dictionary definition of 'about' as "concerning or relating to someone or something; on the subject of them or it" (at [97]).

Giving an example of the Tribunal member's history of car repairs, the AAT stated:

> "A link could be made between the service records and the record kept at reception or other records showing my name and the time at which I had taken the care (sic) in for service. The fact that the information can be traced back to me from the service records or the order form does not, however, change the nature of the information. It is information about the car … or the repairs but not about me" (at [96]).

As a result, the AAT found that mobile network data was about connections between mobile devices, rather than "about an individual", notwithstanding that a known individual - the applicant for access, Mr Grubb - triggered the call or data session which caused the connection. The AAT stated:

> "Once his call or message was transmitted from the first cell that received it from his mobile device, the data that was generated was directed to delivering the call or message to its intended recipient. That data is no longer about Mr Grubb or the fact that he made a call or sent a message or about the number or address to which he sent it. It is not about the content of the call or the message. The data is all about the way in which Telstra delivers the call or the message. That is not about Mr Grubb. It could be said that the mobile network data relates to the way in which Telstra delivers the service or product for which Mr Grubb pays. That does not make the data

[150] *Telstra Corporation Limited and Privacy Commissioner* [2015] AATA 991, 18 December 2015, at [95]; available at http://www8.austlii.edu.au/cgi-bin/viewdoc/au/cases/cth/AATA/2015/991.html

> information about Mr Grubb. It is information about the service it provides to Mr Grubb but not about him" (at [112]).

This interpretation seemed to conflate *object* with *subject*, by suggesting that the primary purpose for which a record was generated is the sole point of reference when determining what that record is 'about'. In other words, the AAT judgment appears to say that what the information is *for* also dictates what the information is *about*. Under such an interpretation, banks could have argued that their records are only 'about' transactions, not the people sending or receiving money as part of those transactions; or hospitals could have claimed that medical records are 'about' clinical procedures, not their patients.

On appeal, the AAT decision was not overturned, with the Federal Court of Australia confirming that there are two elements to the definition of 'personal information': the information must be 'about' an individual, and that individual's identity must be reasonably ascertainable. These can be conceived of as the *about* element, and the *identifiability* element.

The Federal Court stated:

> "in every case it is necessary to consider whether each item of personal information requested, individually or in combination with other items, is about an individual. This will require an evaluative conclusion, depending upon the facts of any individual case, just as a determination of whether the identity can *reasonably* be ascertained will require an evaluative conclusion".[151]

In relation to the *about* element, the Federal Court said:

> "The words 'about an individual' direct attention to the need for the individual to be a subject matter of the information or opinion. This requirement might not be difficult to satisfy. Information and opinions can have multiple subject matters" (at [63]).

By saying that the individual needs to "be a subject matter" of the information, this judgment may have had the effect of slightly narrowing the definition of 'personal information', more so than if the language of "relating to" had been used instead.

However importantly, the Federal Court also said: "even if a single piece of information is not 'about an individual' it might be about the individual when combined with other information" (at [63]).

The judges stressed the need to consider "the totality of the information" (at [63]). In other words, linkability to an identifiable individual might still make something 'personal information'.

[151] *Privacy Commissioner v Telstra Corporation Limited* [2017] FCAFC 4 at [63]; available at https://www.judgments.fedcourt.gov.au/judgments/Judgments/fca/full/2017/2017fcafc0004

The Federal Court made it clear that it was *not* deciding whether or not the metadata to which Ben Grubb was seeking access actually met the definition of 'personal information': "this appeal concerned only a narrow question of statutory interpretation which was whether the words 'about an individual' had any substantive operation. It was not concerned with *when* metadata would be about an individual".

The Federal Court can be seen to have diverged from the AAT's narrower view, by allowing (a) that information may have multiple subject matters, and (b) that the construction of the subject matter can be influenced by the context, i.e. if the data is combined with other data, it might then become 'about' an individual.

The Federal Court decision thus left open the possibility that a piece of data might still be captured by the definition of 'personal information', even though at first glance it appears to have as its subject matter/s not an individual, but something else, such as a telecommunications network, or the environment.

That view is supported by subsequent guidance issued by the OAIC, which states:

> "Information is 'about' an individual where there is a connection between the information and the individual. This is ultimately a question of fact, and will depend on the context and the circumstances of each particular case.
>
> …
>
> Information will also be 'about' someone where it reveals or conveys something about them — even where the person may not, at first, appear to be a subject matter of the information".[152]

The guidance goes on to provide an example of an investigator's report into the cause of the breakdown of a car, which by including a conclusion that 'the owner has contributed to the breakdown by driving negligently' also conveys something about the driver of the car. In such a case, the report includes information about the car *and* about the driver.

The guidance on this point concludes with noting that:

> "Information can have different degrees of connection with an individual and still be personal information. However, at some point, the connection between the information and a person will be too remote for the information to be personal information".[153]

Subsequent to *Privacy Commissioner v Telstra*, the NSW Civil and Administrative Tribunal (NCAT) has noted and followed the Federal Court's view that information and opinions can have multiple subject matters.

[152] Office of the Australian Information Commissioner, *What is personal information?*, May 2017; available at https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information

[153] Office of the Australian Information Commissioner, *What is personal information?*, May 2017; available at https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information

In a case involving the disclosure of an employee's work address to an estranged family member, NCAT found that a work address was "personal information", because it was 'about' the employee, as well as about his employer. The NCAT Member stated:

> "In my view the information is about an individual in that the information was both requested and provided in a context solely concerning the applicant".[154]

NCAT has also rejected an argument made by a respondent that the *Telstra* case meant that information relating to the use of the public transport ticketing Opal Card was 'about' the card, not 'about' an individual:

> "for all relevant purposes, especially concerning a mandatorily registered Gold Opal card, the travel information was more about CNS than about the card. There was no purpose attached to the card information (unique to its requirement for registration) that was not about CNS.
>
> Whilst it is true that the respondent aggregates travel data for planning and other related issues, this is true of all cards, registered or otherwise. On this basis the baseline aggregated Opal travel information attaches to all cards. However, the applicant's card (by being registered) brought that information into the realm of personal information whereby it was not just the number of persons who passed through a transport hub on a certain day etc. It also provided information (where capable of being identified – by the linking to the registered user) which showed the details of the individuals who passed through that hub".[155]

The NSW Privacy Commissioner also supported the argument that "travel history information" is "especially close and relevant" to the individual holding the card, and therefore was personal information about them (at [81]).

[154] *CRP v Department of Family and Community Services* [2017] NSWCATAD 164 at [75]; available at https://www.caselaw.nsw.gov.au/decision/5924eae9e4b058596cba6cad

[155] *Waters v Transport for NSW* [2018] NSWCATAD 40 at [67]-[68]; available at https://www.caselaw.nsw.gov.au/decision/5a8351f1e4b074a7c6e1c492

Case study: Strava fitness data

Strava is a social network of people who like to not only use devices like FitBits to track their movements, heart-rate, calories burned etc, but to then share and compare that data with fellow runners and cyclists. In November 2017 Strava released a data visualisation 'heat map' of one billion 'activities' by people using its app.[156]

In January 2018 an Australian university student pointed out on Twitter that the heat maps could be used to locate sensitive military sites, because military personnel often jog routes just inside the perimeter of their base.[157] Others have noted that the heat map highlighted patterns of road patrols out of military bases in combat zones including in Afghanistan, Iraq, and Syria.[158]

The public release of Strava data highlights a number of issues about the interplay of new technologies, law, and the role of consumer choice and control when using these products and services.

First, the sheer power of geolocation data is incredible. It can show patterns of behaviour of both individuals and groups, and reveal sensitive locations.

Second, geolocation data can be used to find out more about identifiable or already-known individuals. Despite Strava's public position that the heatmap contained only an aggregated and anonymised view of its data, removing identifiers from the data does not make it anonymous. The location data itself can be linked to known individuals who live at a particular address, or matched with other details in the public realm such as social media posts.[159]

Third, privacy harms (including harassment or physical harm) can be done to individuals even if their identity is not revealed. A Strava user has explained how she discovered that her workout routes were accessible to (and commented on by) strangers, even when she thought she had used the privacy settings in the app to prevent public sharing of her data.[160]

Fourth, when individuals comprise a group, say personnel at an army base or worshippers at a mosque or clients of an abortion clinic, the risk posed by or to one becomes a risk for all. While Strava framed the issue as a 'consumer control' matter, pointing users to the app's

[156] See https://medium.com/strava-engineering/the-global-heatmap-now-6x-hotter-23fc01d301de
[157] See https://twitter.com/Nrg8000/status/957318498102865920
[158] Ashley Thomas, "No Place to Hide: Privacy Implications of Geolocation Tracking and Geofencing", *SciTech Lawyer*, Vol 16(2), American Bar Association, 17 January 2020; available at https://www.americanbar.org/groups/science_technology/publications/scitech_lawyer/2020/winter/no-place-hide-privacy-implications-geolocation-tracking-and-geofencing/
[159] Ariel Bogle, "Strava's heatmap revealed military bases, but it also showed nothing is anonymous online", *ABC News Online*, 4 February 2018; available at https://www.abc.net.au/news/science/2018-02-04/strava-heatmap-online-anonymity-is-almost-impossible/9380326
[160] Rosie Spinks, "Using a fitness app taught me the scary truth about why privacy settings are a feminist issue", Quartz, 1 August 2017; available at https://qz.com/1042852/using-a-fitness-app-taught-me-the-scary-truth-about-why-privacy-settings-are-a-feminist-issue/

privacy controls and options to ‘opt out’, a cohort of individuals can only be protected if every single member of that cohort exercises the same degree of consumer control.

Fifth, even at the individual level, privacy controls are difficult for consumers to understand, especially in the context of understanding the bigger risks posed by geolocation data. Many commentators were quick to judge users of the Strava app, saying that military personnel for example should have never allowed themselves to be tracked. Yet a Princeton professor (who happens to be an expert in re-identification) noted that if he couldn’t understand whether or not Strava’s privacy settings actually worked to obscure the user’s home address, or whether the use of a fake name would be enough to prevent cross-linking with other data, a more typical app user cannot be expected to determine where their own level of comfort sits, and how to achieve it.[161]

Finally, privacy controls may not cover all potential uses of personal information, including by the service provider themselves. The Future of Privacy Forum’s review of the app found that Strava’s request for access to location data – the box that pops up on a user’s phone when they first install the app, asking them to ‘Allow’ or ‘Don’t Allow’ – did not mention public sharing. Instead, it said “Allow Strava to access your location while using the app so *you* can track *your* activities” (emphasis added).[162] There was no opportunity for users to opt out of inclusion in a ‘heat map’ likely not even conceived of at the time of initially downloading the app.

[161] Professor Arvind Narayanan, tweet at https://twitter.com/random_walker/status/959885267371995136
[162] Stacey Gray, “If You Can’t Take the Heat Map: Benefits & Risks of Releasing Location Datasets”, Future of Privacy Forum blog, 31 January 2018; available at https://fpf.org/2018/01/31/if-you-cant-take-the-heat-map-benefits-risks-of-releasing-location-datasets/

Case study: NYC taxi trip data

In March 2014, the New York City Taxi & Limousine Commission (TLC) released under FOI data recorded by taxis' GPS systems.[163] The dataset covered more than 173 million individual taxi trips taken in New York City during 2013. The FOI applicant used the data to make a visualisation of a day in the life of a NYC taxi,[164] and published the raw data online for others to use.

Each trip record included the date, location and time of the pickup and drop-off, the fare paid, and any recorded tip. It also included a unique code for each taxi and taxi driver.
In theory the identity of each taxi and taxi driver had been 'anonymised' by the use of 'hashing' – a one-way encryption technique which replaced each driver licence number and taxi medallion number with an alphanumeric code, that can't be reverse-engineered to determine the original information.

However as computer scientist Vijay Pandurangan who found the published dataset pointed out, hashing is not a good solution when you know what the original input *might* have looked like. So given what taxi numbers look like is public knowledge (they are printed on the side of the taxi), he could run the 'hash' against all possible taxi numbers. It took Pandurangan less than an hour to re-identify each vehicle and driver for all 173 million trips.[165]

While the computer science community started debating the limits of hashing and how TLC should have 'properly' de-identified their dataset before releasing it under FOI, postgrad student Anthony Tockar found that even if TLC had removed all the details about the driver and the taxi, the geolocation data alone could potentially identify taxi *passengers*. To demonstrate this, Tockar conducted web searches for images of celebrities getting in or out of taxis in New York during 2013. Using other public data like celebrity gossip blogs, he was able to determine where and when various celebrities got into taxis. Using the TLC dataset, Tockar could then identify exactly where named celebrities such as Bradley Cooper and Jessica Alba went, and how much they paid.[166]

Tockar also developed an interactive map, showing the drop-off address for each taxi trip which had *begun* at a notorious strip club. The same could be done to identify the start or end-point for each taxi trip to or from an abortion clinic, a drug counselling service, or the home address of an investigative journalist, suspected whistle-blower or partner suspected of cheating. As Tockar notes, "with only a small amount of auxiliary knowledge, using this dataset an attacker could identify where an individual went, how much they paid, weekly habits, etc".

[163] Chris Whong, "FOILing NYC's Taxi Trip Data", personal blog, 18 March 2014; available at https://chriswhong.com/open-data/foil_nyc_taxi/

[164] Previously at http://nyctaxi.herokuapp.com/; now available at https://chriswhong.github.io/nyctaxi/ .

[165] Vijay Pandurangan, "On Taxis and Rainbows: Lessons from NYC's improperly anonymized taxi logs", *Medium.com*, 22 June 2014; available at https://tech.vijayp.ca/of-taxis-and-rainbows-f6bc289679a1

[166] Anthony Tockar, "Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset", Neustar blog, 15 September 2014; available at https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/

Case study: Myki public transport data

In mid-2018, Public Transport Victoria (PTV), the agency with responsibility for public transport administration across Victoria, released a dataset of 1.8 billion records of transport users' activity to Data Science Melbourne for use in the Melbourne Datathon. The Datathon is an annual event in which entrants (typically data scientists, academics and students) compete to find innovative uses of a dataset. The dataset contained the records of 'touch on' and 'touch off' activity of 15.1 million myki cards over a three year period to June 2018.

PTV maintained that the dataset was disclosed in response to a request from the Department of Premier and Cabinet (DPC), which oversees the government's open data platform, through the DataVic Access Policy and Guidelines. DPC had been represented on the Datathon judging panel and provided sponsorship to Data Science Melbourne for the Datathon.

Based on advice that certain de-identification techniques would be applied to the data prior to release, PTV completed a threshold Privacy Impact Assessment (PIA) checklist and gave the 'all clear' for the release. However, on their receipt, a number of Datathon competitors reported concerns that the dataset was still readily identifying. Whilst names were excised and card numbers randomised, in a number of cases, taking what might be known from as little as one shared trip with an acquaintance was enough to deduce all trips they had made in the three year period.

A re-identification exercise conducted by a team from the University of Melbourne also found that combining information from other sources with information in the dataset about the relatively small number of some categories of card holder (police and politician card holders), rendered the dataset 'personally identifying'.[167] As a consequence, it revealed a significant amount of location data about individuals and their likely travel patterns.

Due to a number of factors, including the heightened risk posed by the sheer size of the dataset, and the potential impact of the breach on public trust, the Office of the Victorian Information Commissioner (OVIC) determined to investigate the circumstances, including the steps taken to assess and approve the data for release.[168]

The PIA was premised on the assumption that the dataset had been successfully 'anonymised' by one area of PTV, and so concluded that the dataset could therefore be safely released for use in the Datathon.

[167] Dr Chris Culnane, Associate Professor Benjamin I. P. Rubinstein, and Associate Professor Vanessa Teague, "Two data points enough to spot you in open transport records", *Pursuit*, University of Melbourne, 15 August 2019; available at https://pursuit.unimelb.edu.au/articles/two-data-points-enough-to-spot-you-in-open-transport-records

[168] Office of the Victorian Information Commissioner, "Disclosure of myki travel information: Investigation under section 8C(2)(e) of the Privacy and Data Protection Act 2014 (Vic)", 15 August 2019; available at https://ovic.vic.gov.au/wp-content/uploads/2019/08/Report-of-investigation_disclosure-of-myki-travel-information.pdf

For example as to whether the program was going to collect, use or disclose re-identifiable information, PTV's PIA stated:

> "No. There is no way to link the public transport travel patterns of individual mykis to specific people via the encrypted internal card ID – this is not publicly available and will be encrypted in any case. The only remaining risk is that someone may attempt to identify a specific myki card based on the travel patterns but this would require a detailed knowledge of when and where a person had used public transport – basically a travel diary – and it would be very difficult to distinguish from other cards with similar travel patterns. In the unlikely event that this succeeded it would only reveal which Public Transport modes and stops the card had appeared at".

This view, that the dataset had been de-identified, formed the basis for the subsequent governance of the released data. However, as OVIC and the University of Melbourne team demonstrated, re-identification from the dataset was considerably easier than PTV had imagined.

The PIA did not describe in detail what data would be released, other than to say it would be anonymised myki data. The PIA concluded that "no personal information capable of identifying an individual" would be used, but without sufficient analysis or reasoning for that conclusion.

The template did ask users to consider the risk of re-identifiable information and why. But given PTV had firmly concluded that there was no 'personal information' involved, the remaining sections of the template, which were designed to manage risks including that of re-identification, were left incomplete.

OVIC found that PTV had made an unauthorised disclosure and had failed to take reasonable steps to protect the personal information. The agency was thus found in breach of its Data Security and Disclosure obligations.

Individuation was also a factor in this case. The University of Melbourne team also noted that individuals could be singled out and targeted for harm, even if their 'identity' was not knowable. They noted for example that:

> "there are card types for school children of different ages. (The Myki dataset) presents both a security and safety risk by revealing the travel patterns for the card holder over a three-year period: when and where they travel, who they're with and regular times when they're traveling alone".

In other words, a perpetrator could determine where there were young children with a pattern of travelling without an accompanying adult, and those children could be targeted as a result, without needing to know anything about the child's identity.

Glossary

AAT	Administrative Appeals Tribunal
ACCC	Australian Competition and Consumer Commission
APPs	Australian Privacy Principles
CCPA	California Consumer Privacy Act 2018
CDR	Consumer Data Right
DPI	Digital Platforms Inquiry by the ACCC
EU	European Union
GDPR	General Data Protection Regulation 2016
HIPAA	Health Insurance Portability and Accountability Act 1996 (USA)
ICO	Information Commissioner’s Office (UK)
IPC	Information and Privacy Commission (NSW)
NCAT	NSW Civil and Administrative Tribunal
OAIC	Office of the Australian Information Commissioner
OIC	Office of the Information Commissioner (Queensland)
OPC	Office of the Privacy Commissioner (New Zealand)
OPCC	Office of the Privacy Commissioner of Canada
OVIC	Office of the Victorian Information Commissioner
PIA	Privacy Impact Assessment
Privacy Act	Privacy Act 1988 (Cth)
TIA Act	Telecommunications (Interception and Access) Act 1979

Qualifications & confidentiality

The analysis in this report does not constitute legal advice, and should not be construed or relied upon as legal advice by any party. Legal professional privilege does not apply to this report.

This report is made on a confidential basis to our client. It is for the sole discretion of our client to determine whether it will waive confidentiality and provide this report to any other party. The contents of this report will not be divulged to any third party by Salinger Privacy without the express and written permission of our client.

About the author

This report has been prepared by Anna Johnston, Principal, Salinger Privacy.

Ms Johnston was previously the Deputy Privacy Commissioner of NSW. She holds a first class honours degree in Law, a Masters of Public Policy with honours, a Graduate Certificate in Management, a Graduate Diploma of Legal Practice, and a Bachelor of Arts. Ms Johnston was admitted as a Solicitor of the Supreme Court of NSW in 1996. She established Salinger Privacy in 2004.

Anna has been called upon to provide expert testimony before various Parliamentary inquiries and the Productivity Commission, spoken at numerous conferences, and is regularly asked to comment on privacy issues in the media. She is a lifetime member of the Australian Privacy Foundation, a member of the International Association of Privacy Professionals (IAPP) since 2008, and in 2019 was recognised as an industry veteran by the IAPP with the designation of Fellow of Information Privacy (FIP). She is also a Certified Information Privacy Professional, Europe (CIPP/E) and a Certified Information Privacy Manager (CIPM).

Anna’s *pro bono* advisory, advocacy, academic and editorial positions have included:

- Visiting Scholar at the Research Group on Law, Science, Technology and Society of the Faculty of Law and Criminology of the Vrije Universiteit Brussel (VUB), 2018
- Advisory Board Member for the EU’s STAR project, to develop training on behalf of European Data Protection Authorities, 2017 to date
- Member of the Asian Privacy Scholars Network (APSN), 2017 to date
- Editorial Board Member, Privacy Law Bulletin, 2010-16
- Advisory Committee Member, the Australian Law Reform Commission’s Inquiry into the Invasion of Privacy, 2013-14
- Board member, the International Association of Privacy Professionals, Australia & New Zealand, 2010-11
- Advisory Committee Member, the Australian Law Reform Commission’s expert advisory group on health privacy for the Review of the Privacy Act, 2006-08
- Project Team Member, University of New South Wales’ Interpreting Privacy Principles project, 2006-09
- Campaign Director, NoIDCard, the successful campaign against the proposed Access Card, 2006-07
- Chair of the Australian Privacy Foundation, 2005-06; Member of the International Committee 2018 to date
- Consumer Representative Member, National E-Health Transition Authority’s Consumer & Clinician Forum, 2005-06
- Primary Author, Australian country reports for Privacy International’s Privacy and Human Rights, 2005-06
- Editorial Board Member, Privacy Law & Policy Reporter, 2004-06

About Salinger Privacy

Salinger Privacy offers specialist consulting services on privacy and data governance matters, including Privacy Impact Assessments and privacy audits, and the development of privacy-related policies and procedures.

Salinger Privacy also offers a range of privacy guidance publications, e-learning and face-to-face compliance training options, and Privacy Tools such as templates and checklists.

SalingerPrivacy

We know privacy inside out.

Salinger Consulting Pty Ltd
ABN 84 110 386 537
PO Box 1250, Manly NSW 1655
www.salingerprivacy.com.au