Data – it's not just about breaches

Patrick Sefton, Principal, Brightline Lawyers
11th Annual In-House Counsel Conference: Innovate or Perish, 13 March 2018

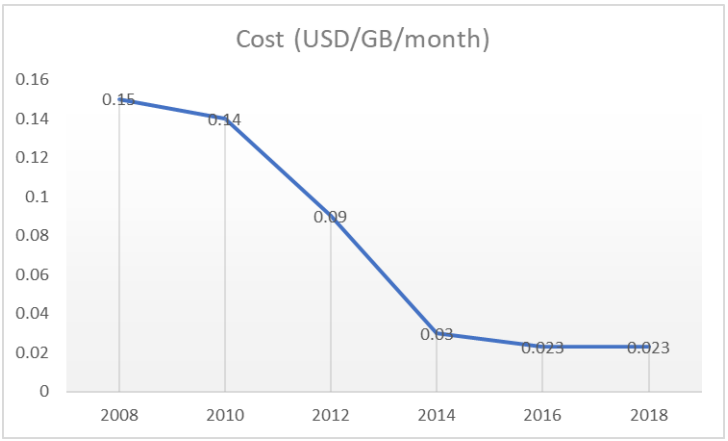
Data has long been a problem child for the legal and accounting professions. Although this paper is not written from an accounting perspective, it seems that accounting tends to group business intangibles, perhaps excluding formally protected IP rights, as "goodwill" or similar and nether separately account for or value them within an enterprise. Similarly, the law has been disinterested, or deliberately averse, to recognising or protecting information or data, as distinct from formally legislated classes of IP rights.

Meanwhile, unburdened by meaningful legal or accounting scrutiny, the nerds[1] have been busy…

There is a *lot* of data, and creation and storage is on an exponential path. It is driven by a number of factors including:

- cloud storage cost and accessibility;
- ICT decision-making to store marginal data that would previously have been discarded;
- organisational paranoia resulting in collection and storage "just in case";
- some far-sighted organisations implementing or planning Machine Learning AI, which improves broadly speaking in proportion to the quantity of training data available.

On the first of these drivers, cost, over the last 10 years the cost of AWS cloud storage has fallen by roughly 90% in real terms:



AWS Cloud Data Storage Cost 2008-2018

At the same time, accessibility and ease of use for cloud data has improved, and providers are now bringing on-line an array of data processing tools, for example in the fields of efficient queries, business analytics and data warehousing, for use with cloud repositories.

Lurking in the background are Machine Learning (ML) type Artificial Intelligence systems. ML is essentially learning by pattern-matching at complexity and scale, and underlies many other advances such as automated language translation, voice recognition and image classification. ML is becoming a significant transformational technology in which large volumes of training data are key to high-

[1] With respect.

performing systems. Predictably, cloud services providers are developing accessible, packaged and commoditised ML frameworks for use with (cloud-based) client training data.

The operational details of ML and opportunities for the legal profession are beyond the scope of this paper. It is mentioned here by way of one explanation for the continuing exponential growth of data creation and storage, both in preparation for the use of such systems, and because it is understood that "more data is better" for their performance. Some of the legal risks inherent in ML systems are also mentioned in the 'leveraging data' section below.

Data as IP: Who owns what?

Informally, data and information is described in proprietary terms. Formally, however, the law is uncomfortable with ownership of information and data *per se*, aside from formal IP regimes such as copyright, and from the intervention of equity to protect confidential information.

Is data "property"?

In Queensland, Schedule 1 to the *Acts Interpretation Act*[2] defines "property" as:

> "any legal or equitable estate or interest (whether present or future, vested or contingent, or tangible or intangible) in real or personal property of any description (including money), and includes things in action."

It's a self-referential definition, but confirms that references to "property," at least in Queensland legislation, have a broad interpretation, including personal property and things in action such as legal rights.

However the courts have consistently declined to describe information or data as property.

In the now quaintly primitive *Victoria Park Racing v Taylor*[3] the High Court confirmed in relation to unauthorised race calling from a site adjacent to a racecourse:

> The defendant does no wrong to the plaintiff by looking at what takes place on the plaintiff's land. Further, he does no wrong to the plaintiff by describing to other persons, to as wide an audience as he can obtain, what takes place on the plaintiff's ground. […]
>
> A "spectacle" cannot be "owned" in any ordinary sense of that word. Even if there were any legal principle which prevented one person from gaining an advantage for himself or causing damage to another by describing a spectacle produced by that other person, the rights of the latter person could be described as property only in a metaphorical sense.[4]

In the more contemporary *Breen v Williams*,[5] a case concerning a patient's access to medical records created by their treating medical practitioner, the High Court confirmed "in general, information is not property at all"[6] (subject to equity intervening to prevent disclosure of confidences). The records (that

---

[2] *Acts Interpretation Act 1954* (Qld).
[3] *Victoria Park Racing & Recreation Grounds Co Ltd v Taylor* [1937] HCA 45; (1937) 58 CLR 479 (26 August 1937).
[4] *Victoria Park Racing v Taylor* per Latham CJ.
[5] *Breen v Williams* ('the medical records access case") [1996] HCA 57; (1996) 186 CLR 71 (6 September 1996).
[6] *Breen v Williams* per Brennan J at [12], citing *Phipps v Boardman* [1966] UKHL 2; (1967) 2 AC 46 at 127-128.

is, their physical representation as chattels) belong to the doctor, and the copyright in them does too. If there is no specific legal obligation for access, the owner is entitled to refuse to produce them by virtue of ownership, and refuse to allow copies to be made (by virtue of copyright). The *information* itself is not property and not capable of ownership.

Two more examples from recent caselaw illustrate some of the consequences of data and information not being considered property.

Your Response v Datateam Business Media[7] concerned a dispute about access to a database.

Datateam was a magazine publisher which maintained a database of subscribers. Your Response (YR) was a database manager whose role was to hold, amend and update its clients' databases.

Datateam and YR entered into a contract described by the court as 'vague' and evidenced by email and some discussions, under which YR would maintain and update Datateam's subscription database. The contract was silent on provision of the updated database by YR back to Datateam, on termination of the contract, and what the consequences would for the database on termination.

The relationship deteriorated. Ultimately Datateam purported to terminate. YR sent an invoice to Datateam. Datateam requested the database be returned, but YR refused to provide it. YR ceased providing services and refused to release the updated database or provide access to it until all outstanding fees were paid by Datateam. Datateam refused to pay until the database was made available to it. YR sued for its fees, and for damages for repudiation of the contract. Datateam counterclaimed for damages for the cost of reconstructing the database.

Among the issues at trial were, could YR exercise a common law possessory lien over the database pending payment of its outstanding fees?

The Court reviewed the caselaw on possessory liens, for example, bookkeeper's lien over ledgers, solicitor's over client files. The Court distinguished tangible from intangible property including things in action. The Court considered proprietary torts such as conversion and detenue, which only relate to property and not other subject matter. The Court also considered that possession or transfer of possession of the database *per se* was not meaningful: the item actually possessed was only the tangible media on which the database was recorded, and that "practical control" (of the database) is not the same as "possession" (of a chattel).

Ultimately, the Court followed the existing caselaw that liens, like proprietary torts such as conversion, are founded in possession of tangible property and therefore should not extend to intangibles or quasi-property. The Court was quite transparent in its reasoning as to whether it should extend the concept, but ultimately decided that it was open for parties to deal with the issue by contract – the court expressly resisted YR's "appeal to modernity" because it considered there may be unintended consequences, for example, an impact on creditors in insolvency, on secured lenders, or on other law such as larceny. The court therefore held that YR was not entitled to exercise a common law possessory lien in respect of the database.

[7] *Your Response Ltd v Datateam Business Media Ltd* [2014] EWCA Civ 281 (UK Court of Appeal).

There was a similar argument with a similar outcome in *Dixon v R*.[8] Mr Dixon worked as a bouncer at the Altitude Bar in Queenstown during the 2011 Rugby World Cup. He obtained from the bar's security systems video footage of the captain of the English team "socialising" with a female companion on the dance floor. After unsuccessfully attempting to sell the footage, Mr Dixon placed the video on YouTube. This resulted in significant publicity, as the player in question was in fact married to a member of the British royal family.

Mr Dixon was charged with dishonestly *obtaining property* by accessing a computer system.[9] The question raised in the Court was whether the video recording was "property" within the NZ Crimes Act. The definition under the Act includes things like "money" and "electricity" but not a broader expression of information or other intangibles.

The Court decided the recording was not "property" within the Act, saying:

> "…the courts have essentially taken the view that any illogicality is outweighed by the strong policy reasons that militate against recognition of information (whether confidential or otherwise) as property. The concern is that if the law were to recognise confidential information as property and so afford it the full protection of property law, that would be likely to have a damaging effect on the free flow of information and freedom of speech."[10]

(Unfortunately for Mr Dixon, the Court substituted an alternative verdict that he had dishonestly obtained a benefit, rather than property, from the recording, so he was in no better position.)

Summary – data as property

In summary, pure information (whether confidential or otherwise) is not property. This may have positive or negative outcomes, depending on circumstances.

Intellectual property rights, however, can certainly be property (such rights are in most instances a thing in action: a right to sue, but are also expressly stated by statute to be property). For example, s196(1) of the Copyright Act[11] provides "Copyright is personal property."

Copyright

Copyright is the principal legal mechanic for protecting written work. It expressly protects software (defined as "a set of statements or instructions to be used directly or indirectly in a computer in order to bring about a certain result."[12]). It confers rights one would usually expect as an "owner" of written work, that is, the exclusive right to control reproduction, transmission, publication and adaptation of the work.

However, copyright does not necessarily apply to data, even where there has been significant effort and cost in development of a particular dataset. Inclusion now turns on whether there has been sufficient originality of authorship.

---

[8] *Dixon v R* [2014] NZCA 329; [2014] 3 NZLR 504; (2014) 27 CRNZ 129 (Court of Appeal of New Zealand, 17 July 2014).
[9] s249(1)(a) of the *Crimes Act 1961* (NZ).
[10] *Dixon v R* at [34] per French J.
[11] *Copyright Act 1968* (Cth).
[12] *Copyright Act 1968* (Cth) s10.

IceTV v Nine[13] was copyright litigation concerning the publication of an electronic program guide for use as a guide to broadcast TV scheduling. Essentially IceTV reproduced the scheduling guides published by individual television channels, including Nine's. Ultimately the court decided that there was no copyright in Nine's published guide, as there was insufficient creative originality in the elements of Nine's guide that were copied. "Copyright does not protect facts or information."[14]

Similarly, in *Telstra v Phone Directories Company*[15] the Full Federal Court held that the white and yellow pages, which are undisputedly compilations of factual information, and the compilation of which is partly automated, do not attract copyright protection. Australian copyright law does not protect the skill and labour in compiling material, but in original creative authorship.

Both IceTV and PDC are authoritative caselaw to the effect that there is no copyright in collections of data where there has been no creative authorship. This implies that there is no copyright protection for most databases, particularly databases that have been automated or partly automated in their development. The subsistence of copyright requires significant human authorship by way of original intellectual effort.

Confidential Information

In the course of describing information as outside the law of property, the courts have consistently also noted that information the subject of a confidentiality obligation is a special case.[16] Confidentiality can therefore be applied to protect the exclusivity in data that is genuinely confidential.

Confidentiality obligations can be imposed by contract and by an equitable duty in appropriate circumstances. The contractual obligation will turn on the terms of the contract. The equitable obligation arises where two criteria are met:[17] that the information is in fact confidential, and that it has been imparted in circumstances importing an obligation of confidence, that is, where the recipient knew the information was confidential or should have realised from the circumstances of the disclosure.

The criteria can practically be applied in a contemporary context to data as follows.

Criterion	**Organisational response to demonstrate in respect of data**
Information is in fact confidential	• Demonstrate data is treated confidentially; • Data is not public(!) • Identification of confidential data; • Internal communication as to status of data as confidential; • Internal access controls;

[13] *IceTV Pty Limited v Nine Network Australia Pty Limited* [2009] HCA 14 (22 April 2009).
[14] *IceTV v Nine* at [28] per French CJ and Crennan and Kiefel JJ.
[15] *Telstra Corporation Limited v Phone Directories Company Pty Ltd* [2010] FCAFC 149 (15 December 2010).
[16] For example, *Breen v Williams* at [12] per Brennan J.
[17] *Coco v A N Clark (Engineers) Ltd* [1969] RPC 41 per Megarry J at [47]; *Corrs Pavey Whiting & Byrne v Collector of Customs (Vic)* [1987] FCA 266 (13 August 1987) per Gummow J at [14].

Criterion	Organisational response to demonstrate in respect of data
	• Practical external controls, eg, access to only part of database, access for query only, throttles on queries or access; • Access only available to individuals with need to know/access; • Audit trail and audit to confirm controls operational and effective.
Information has been imparted in circumstances of confidence	• Confidentiality notice or click-wrap to recipients; • Precedent text in email or other communications under which confidential information is provided; • Screen or printed copy annotations of "confidential" or "commercial-in-confidence"; • NDA or other contractual non-disclosure terms.

Leveraging data –
How can, and should, we use data and information?
Potential pitfalls and perils in data use

In July 2017, Wetherspoons, the 1,000 venue, $3B turnover UK hotel chain, announced the deliberate and permanent deletion of its entire email marketing database.[18] This was arguably the result of a number of factors including the pending commencement of stricter European information privacy controls known as the General Data Protection Regulation (GDPR), and mitigation of the fact that Wetherspoons had been the subject of breaches in the past, but ultimately it was a decision that, on balance, the value of the data for marketing and customer contact purposes was outweighed by regulatory compliance cost and breach risk.

Keeping and leveraging data in an organisation involves a number of trade-offs. As the Wetherspoons example shows, perhaps the most significant involves whether particular data should be retained at all, or for any significant period of time beyond its immediate usefulness.

The opposing ends of the spectrum of approaches may be summarised as follows:

"Keep only the essentials, and then only while useful"	"Keep everything, forever"
• Reduce risk of data breach (disclosable or otherwise); • Minimise "attack surface," minimise attractiveness as breach target; • Minimise compliance costs;	• storage and processing are now very, very, low cost; • you never know when something might be useful later; • [special case of previous point] Machine Learning AI is becoming an important transformative

[18] "Wetherspoons just deleted its entire customer email database – on purpose", *Wired*, 3 July 2017, <http://www.wired.co.uk/article/wetherspoons-email-database-gdpr>.

"Keep only the essentials, and then only while useful"	**"Keep everything, forever"**
• [personal information only] compliance with APP 3.1/3.2: [entity] must not collect unless reasonably necessary for function or activity; • [personal information only] compliance with APP 11.2: if [entity] holds personal information … and no-longer needs the information for any [lawful] purpose … and not *required* to retain … then *must* destroy or de-identify.	technology, and large volumes of training data is key to leveraging this technology; • [private sector] data increases value as organisation; • [public sector] data improves decision-making, policy-setting capacity.

Ultimately, and subject to compliance obligations, it is a decision for each organisation where to place themselves on this spectrum. Clearly, an organisation may take a different approach in respect of personal information or other sensitive data, as compared with non-sensitive or appropriately de-identified data.

But where an organisation has decided to keep significant datasets, the question becomes how to derive meaningful insight from the organisation's data. There may be a need to overcome inertia in the sense of changing the business-as-usual approach because of a data-driven insight. There will be a need to change decision-making from a heuristic or even "gut-feel," to an objective, data-driven approach. Such organisations must become more open to new ideas, and more learning-oriented than received-wisdom.

Before becoming an entirely data-driven organisation, there are, however, some emerging warnings about data quality and bias.

There is, for example, a growing body of criticism about whether Machine Learning AI is entrenching discrimination & bias. For example: facial recognition systems unable to recognise coloured or Asian faces. Machine Learning is sometimes considered a mirror (because it processes and pattern-matches real-life data), and its users and proponents are sometimes disappointed in what they see. Machine Learning does not understand context nor ethical consequences (nor legal obligations!). For example, if there are fewer minorities represented in an ML system's training data (by definition) this therefore results in predictably poorer decision-making in respect of such minorities as a group. In addition, a history of bias in data will be replicated in ML decisions.

To take an example: several well-known public corpuses of image recognition training images have (predictable) gender bias. So, activities such as people cooking are disproportionately female, and people playing sports disproportionately male. ML AIs trained on that data inherit and amplify the bias.[19]

Discrimination and bias in ML systems can be particularly problematic in applications such as hiring, healthcare, administrative decision-making, criminal justice.

For example, the PSA (Public Safety Assessment) pretrial risk assessment tool in the California criminal justice system is used as one aspect of risk assessment for release of accused on bail. However the system is opaque and has been the subject of non-disclosure obligations from its providers. Overall, the system has been "good" in that more individuals were released on bail, and

[19] "Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints," J Zhao, T Wang, M Yatskar, V Ordonez, K-W Chang,

overall recidivism of those on bail has been reduced. But the system is impenetrable in individual cases, there is no information about its 750,000-case training dataset and no information about validation testing.

Privacy considerations of de-identification and implications for big data usage

Re-identification risk

In August 2016 the Commonwealth Department of Health published a dataset containing longitudinal (1984–2014) medical billing records of one tenth of all Australians with Medicare cards – approximately 2.9 million individuals.[20] The dataset comprised about a billion rows, each row being an individual Medicare or PBS claim transaction. The release was done without the consent of the individuals concerned, nor any restrictions on access, on the basis that the data was de-identified and therefore no-longer "personal information" within the Privacy Act, because individuals could no-longer be identified within it. For each row, the Department published an "encrypted" patient ID, year of birth, and gender, together with the transaction details (for example, claim type, such as GP consultation or pregnancy management). They also removed unique or unusual events which might be easily matched with public information, and perturbed each event date by up to two weeks, to reduce the risk of re-identification by matching-up specific details of events.

The publication was undertaken with good intentions – the data was proposed to be used for medical research that reduces duplication and costs, ultimately improving health outcomes.[21]

Unfortunately, the Department's approach to de-identification was badly flawed. A team from the University of Melbourne quickly found that provider (and impliedly also patient) identification numbers could be entirely re-identified through a cryptographic attack on a poorly designed encryption scheme.[22] The team also found that patients could be re-identified through a process of linking the published data with known information about individuals such as known medical procedures (for example, childbirths, sports injuries) and the patient's year of birth (published as part of the data). The University of Melbourne team found unique matches in the data for high-profile individuals (for example, an AFL player who was known to have had a particular surgery on a particular date).[23]

More broadly, the University of Melbourne team also made some more general comments about the nature and limits of de-identification, concluding that:

- it's a technically challenging task to understand whether a particular algorithm securely encrypts data or not, and expert assistance should be sought;

- datasets containing sensitive information about individuals clearly deserve more caution than others, and may not always be suitable for open public release;

---

[20] "Govt releases billion-line 'de-identified' health dataset," *IT News*, 15 August 2016, <https://www.itnews.com.au/news/govt-releases-billion-line-de-identified-health-dataset-433814>.
[21] "How the govt is using Medicare data to improve the health system," IT News, 4 December 2015, <https://www.itnews.com.au/news/how-the-govt-is-using-medicare-data-to-improve-the-health-system-412690>.
[22] "Health Data in an Open World", C Culnane, B Rubinstein and V Teague, School of Computing and Information Systems, The University of Melbourne, 18 December 2017, <https://arxiv.org/ftp/arxiv/papers/1712/1712.05627.pdf>.
[23] Above, note 22, at p10.

- compliance with de-identification guidelines, however well-intentioned, cannot be assumed to guarantee privacy protection;

- a dataset of the size and detail of the MBS/PBS 10% sample cannot be published in a useful form while retaining individual privacy.

The University of Melbourne team's paper[24] describing their analysis of the release is quite accessible, and quite sobering.

The Department subsequently suspended access to the data. But of course, once data is published, it cannot be "unpublished," and copies may have been taken. The dataset may therefore be vulnerable far into the future, and be subject both to more sophisticated cryptographic attacks and to linkage attacks from additional data sources.

A further Australian government response to this event was to propose the criminalisation of data re-identification (for de-identified data released by Commonwealth agencies). The *Privacy Amendment (Re-identification Offence) Bill 2016* (Cth) was introduced in October 2016 and resulted in conflicting committee responses from the Senate Legal and Constitutional Affairs Legislation Committee in February 2017. The Bill has not progressed.

The new De-Identification Decision-Making Framework (DDF)

The OAIC, together with CSIRO's data science organisation Data61, have adapted and published an Australian version of a UK de-identification guide, known as the *De-Identification Decision-Making Framework*.[25]

The Framework is a comprehensive but non-technical approach to issues raised in data de-identification. The Framework is not necessarily a "how to" but can be considered a general model of issues that arise in a de-identification project. In a significant project, it would need to be supplemented by expert advice, but in that case can be used as a basis for seeking and scoping such assistance.

The Framework comprises ten components within three core activities:

Situation Audit (identify and frame issues, understand relationship between data and its environment, scope process, clarify goals)

1 Describe data situation (relationship between data and environment)
2 Understand legal responsibilities (is data personal or de-identified? if de-identified, how to maintain that status? data ownership? contractual or other controls on use?)
3 Know the data (subjects, data type, dataset properties eg age/quality, sensitivity)
4 Understand the use case (reason for share/release, who will access, how accessed)
5 Meet ethical obligations (consent if practical, or transparency, engagement, governance)

Risk Analysis and Control (technical processes)

6 Identify processes to assess disclosure risk (difficult, requires judgement, good structure and detailed guidance in framework)
7 Identify disclosure control processes relevant to data situation (the environment, ie, who has access, how and for what purposes, and the data, ie, aggregation and/or perturbation)

[24] Above, note 22.
[25] <http://data61.csiro.au/en/Our-Work/Safety-and-Security/Privacy-Preservation/De-identification-Decision-Making-Framework>.

Impact Management (communication and planning)

8 Identify stakeholders (plan communication and engagement)

9 Plan next steps after share/release (keep register, maintain awareness of environment)

10 Plan what if things go wrong (documentation and audit trail, crisis management plan)

The Framework places emphases on consideration of data in context (the "data situation"), and is a useful resource for identifying issues and for a structured approach to a significant data release.

The Framework makes clear that there will still be complex judgement calls. There may also be situations where it is not possible to release detailed (de-identified) data and to preserve privacy.

Data breach obligations and other areas of potential liability

The notifiable data breach ("NDB") provisions in the new Part IIIC of the Privacy Act[26] came into effect on 22 February 2018. There has been broad (and in some instances somewhat alarmist) commentary concerning the impact of and potential exposure arising from the NDB provisions. Consequently this paper will not traverse the obligations in the data breach provisions in detail. Instead we will consider some of the exclusions and subjectivity present in the legislation, from the perspective that the legislation is more "light-touch" than might be apparent from some of the commentary.

It is well-understood that the NDB provisions will not generally apply to "small business," being business with less than $3M annual turnover.[27] The small business exemption is said to exclude some 94% of Australian enterprises from the scheme.[28]

The NDB provisions also do not in general apply to state government departments and agencies. In Queensland the *Information Privacy Act*[29] continues to apply, and that legislation does require attention to data security, but does not mandate breach disclosures. Of course, affected departments and agencies may still disclose. The Information Commissioner Queensland supports actions that "allow affected individuals to take control of their personal information."[30] (It should also be noted that many state government entities will be tax file number recipients[31] and therefore subject to the NDB provisions in respect of tax file number information.[32])

The NDB scheme applies only to *unauthorised* access or disclosure.[33] So an intentional (even if misguided) disclosure of personal information is not regulated by the NDB provisions. This would include, for example, the MBS 10% disclosure discussed above, where purportedly de-identified personal information could be readily re-identified.

---

[26] Part IIIC – Notification of Eligible Data Breaches, *Privacy Act 1988* (Cth).
[27] *Privacy Act* s6D.
[28] Explanatory Memorandum to the *Privacy Amendment (Notifiable Data Breaches) Bill 2016* (Cth) at [184].
[29] *Information Privacy Act 2009* (Qld)
[30] OIC Queensland News Releases, "Privacy breach notification," 20 February 2017, <https://www.oic.qld.gov.au/about/news/privacy-breach-notification> and "Updated guidance on privacy breach management," 12 February 2018, <https://www.oic.qld.gov.au/about/news/updated-guidance-on-privacy-breach-management>.
[31] Refer *Privacy Act* s11.
[32] *Privacy Act* s26WE(1)(d).
[33] Privacy Act s26WE(2).

The NDB scheme also applies only where the disclosure is *objectively likely to result in serious harm*.[34] (That is, the harm must be both likely to occur and serious in kind). There is some guidance in the legislation[35] about what this means (discussed below). The explanatory memorandum[36] suggests that serious financial, economic or physical harm would be relevant risks, and that mere personal distress or upset would not be sufficient.

The guidance in the legislation around serious harm and likelihood could be interpreted as a set of de facto exceptions. For example, whether the information is protected by security measures[37] such as encryption invites the conclusion, if encryption is present, that the risk of serious harm is unlikely. This is so, even though security measures can arguably provide a false sense of security. Similarly, a conclusion that only the general public, and not a motivated attacker, comprises the class of persons who could obtain the information[38] again invites a conclusion that the risk of serious harm is unlikely.

A further exception applies where an entity takes remedial action before any serious harm has occurred.[39] An entity can also obtain up to 30 days' grace to comply if it considers it only has a suspicion, not knowledge, that an eligible breach has occurred.

In relation to disclosure and notification itself, note there is an option, if it is not practicable to notify individuals concerned, to notify by placing the notification on the entity's own website and taking "reasonable steps to publicise."[40] There is also an exception if another entity has already disclosed and notified the same breach:[41] potentially incentivising delay. Finally, the OAIC may declare that disclosure and notification are not required in respect of a particular breach[42] or may extend the time for notification.[43] Significantly, if an entity applies for such a declaration, the entity is not required to disclose or notify the breach until the OAIC makes a decision on the application.[44] The significance of this provision is brought into further relief when it is noted than the OAIC has received no additional funding to administer the NDB scheme, and has admitted it already experiences delay over privacy complaints in other areas.[45] The OAIC has indicated it will take a priority approach to all of its work.[46] The OAIC has also indicated it will not be publishing the names of organisations that disclose data breaches, at least for the first 12 months of operation of the scheme.[47]

Remedies for contravention of the scheme, read in context, are also comparatively light-touch. The civil penalty regime is only applicable where there is a "serious or repeated" interference with privacy.[48] OAIC guidance concerning penalties is that even where a penalty approach would be open under the legislation, the office will only consider them in cases of particular seriousness or egregious conduct, and take into account the compliance history of the organisation, any failure to take

---

[34] Privacy Act s26WE(2).
[35] Privacy Act s26WG.
[36] Explanatory Memorandum at [8]-[9].
[37] Privacy Act s26WG(e).
[38] Privacy Act s26WG(g).
[39] Privacy Act s26WF.
[40] Privacy Act s26wl(2)(c).
[41] Privacy Act s26WM.
[42] Privacy Act s26WQ(1)(c).
[43] Privacy Act s26WQ(1)(d).
[44] Privacy Act s26WQ(9).
[45] "Privacy Commissioner's small budget to make policing new data breach laws difficult, experts say," *The Sydney Morning Herald*, 23 February 2018, <https://www.smh.com.au/technology/privacy-commissioner-s-small-budget-to-make-policing-new-data-breach-laws-difficult-experts-say-20180223-p4z1dj.html>.
[46] Above, note 45.
[47] Above, note 45.
[48] Privacy Act s13G.

obligations seriously, or any blatant disregard for the regulation.[49] It seems to follow that organisations which make good faith efforts to comply are remotely unlikely to be formally penalised.

Other areas of liability – class actions

Because, by their nature, large numbers of individuals may be affected by a data breach, class action lawyers and funders are active in this area. In the US, more than 50 actions are current against Equifax in respect of its July 2017 data breach. Typical allegations are of negligence, breach of statutory duty, and breach of consumer law.

In respect of its June 2017 incident, Equifax has announced costs of USD439M to date (USD125M of which will be covered by its cybersecurity insurer) and has suggested the total cost could exceed USD600M.[50] The costs arise from technology and security upgrades, legal fees and free credit monitoring services for individuals affected.

In Australia there is at least one current class action in respect of a data breach. In November 2017 an action was commenced on behalf of NSW Ambulance employees and contractors in respect of a 2013 data breach caused by a rogue contractor who allegedly exfiltrated and sold employee data to third parties.[51] The claim alleges breach of confidence, breach of contract, misleading or deceptive conduct, and tortious invasion of privacy.

Commercial transactions involving data

Data is a significant business asset. It should be specifically accounted for in commercial transactions. Such transactions may include exploitation for value of the data itself, or broader corporate transactions which may involve data as one aspect.

Initial considerations

Three initial considerations for transactions involving data are whether the data is:

- subject to copyright;[52]
- genuinely confidential;[53] and/or
- regulated as "personal information"[54] or otherwise.

As indicated earlier in this paper, subsistence of copyright in a dataset turns on whether there has been a sufficient degree of original creative authorship. Most datasets of factual (in the copyright sense,

---

[49] OAIC Guide to Privacy Regulatory Action, Chapter 6: Civil penalties — serious or repeated interference with privacy and other penalty provisions, at [6.27], <https://www.oaic.gov.au/about-us/our-regulatory-approach/guide-to-privacy-regulatory-action/chapter-6-civil-penalties>.
[50] "Equifax breach could be most expensive hack in history," *CRN*, 5 March 2018, <https://www.crn.com.au/news/equifax-breach-could-be-most-expensive-hack-in-history-486256>.
[51] "Paramedics launch class action over the sale of their medical records to personal injury solicitors," The Sydney Morning Herald, 18 November 2017, <https://www.smh.com.au/national/nsw/paramedics-launch-class-action-over-the-sale-of-their-medical-records-to-personal-injury-solicitors-20171118-gzo44u.html>.
[52] *Copyright Act 1968* (Cth) s32.
[53] *Re Corrs Pavey Whiting and Byrne v Collector of Customs of Victoria and Alphapharm Pty Ltd* [1987] FCA 266 (13 August 1987) per Gummow J at [14].
[54] *Privacy Act 1988* (Cth) s6.

"non-original") data will not attract copyright. Similarly, copyright will not subsist in most automatically generated data.

Also as earlier indicated, evidence of steps taken to protect and maintain the confidentiality of a dataset would be required to qualify the data for protection as confidential information. The confidential data must be identified "with specificity, and not merely in global terms" and the owner must prove the data "has the necessary quality of confidentiality."

In relation to the regulation of data as personal information, after *Privacy Commissioner v Telstra*[55] it is clear that "personal information," with particular reference to the requirement in the definition that the data be "about an identified individual," is limited to information where the individual is the subject matter of the data, not merely information about some other subject matter, even if it relates to the individual or if from it the individual may still be identified.

Treatment of datasets in transactions

The above threshold classification questions may lead to different approaches to the treatment of datasets in transactions:

- Datasets that are **copyright** may be more robustly published, since the proprietary and effective international nature of copyright allows exclusivity to be enforced against third party infringers.

  Relevant due diligence enquiries therefore go to subsistence and the providence of ownership, and the nature of any existing licence arrangements.

- Data that is said to be **confidential** is more vulnerable, as the right is effectively enforceable only against particular class of recipients (where disclosure occurred in circumstances of confidence, or where subject to a contractual confidentiality obligation).

  Relevant enquiries include the steps taken to maintain the confidentiality of the data (relevant both practically and legally), and the nature of the relationships in which the data has been disclosed, including the terms of any confidentiality agreements.

- Data which includes **personal information** carries a regulatory burden, as the data will usually be subject to privacy compliance obligations.

  Enquiries about privacy compliance include details of compliance process, the terms of any consents from relevant individuals, whether the data is used or disclosed outside the relevant APP or consent, and relevant information security approach. The last issue is, of course, relevant to the now elevated consequences of security breach. Regulatory intervention in the field is rare, but of course should also be investigated and disclosed, as should any inquiries or complaints from individuals.

[55] *Privacy Commissioner v Telstra Corporation Limited* [2017] FCAFC 4 (19 January 2017) at [63] per Kenny and Edelman JJ.

Commercial exploitation of data

Data is typically externally commercially exploited via licence agreements. "Licence" terminology is conventional, notwithstanding that copyright often does not subsist, and therefore technically a licence is not required – non-copyright data provision and usage is governed by contract only. Licences to use data may be described and confined in the usual ways: exclusive, sole or non-exclusive, geographical limits, time period limits, and field- or purpose-of-use. Other aspects to be included in a data licence include:

- if the data is confidential, confidentiality terms;
- if the data contains personal information, obligations to comply with privacy regulation and to assist with individual and regulatory enquiries and complaints;
- an obligation to deliver-up the data, including any modifications, and to delete and not retain any copies, on termination or expiry (and also periodically or on demand, if appropriate);
- warranty and liability terms, typically excluding (and possibly indemnifying from) liability arising from use of the data;

Further note:

- dealing in personal information (that is, buying or selling personal information in the course of a business) means an entity cannot take advantage of the "small business operator" exemption in the Privacy Act.[56]

Data can also of course be exploited by sale, whether a stand-alone sale of the date or, more likely, in the course of an entity or business sale. In relation to data included in a sale, a purchaser:

- will likely investigate the provenance of the data, including an audit or other expert assessment to determine whether the data is as described. A purchaser might also seek an expert valuation of the data;
- may seek assurances as to its exclusivity in the data, whether by way of assurances as to the maintenance of its confidentiality, or confirmation of its copyright status;
- will need to consider the regulatory burden of data that includes personal information, and obtain disclosure and warranties concerning regulatory compliance.

Data loss as excluded consequential loss

Finally on commercial transactions involving data, and in particular in relation to liability issues, it is becoming common to include "loss of data" as an element of "indirect" or "consequential" loss which is routinely excluded from recovery by parties to all kinds of commercial agreements, but particularly provision of technology services.

[56] Privacy Act 1988 (Cth) s6D(4)(c) and (d).

Loss of or damage to a valuable business asset is hardly “consequential loss,” in the sense that term is normally used, that is, as a category of unrecoverable intangible or indirect, purely economic losses such as loss of anticipated revenue.

Nevertheless its common inclusion in this category is, presumably, based on the premises that the amount of the loss may be disproportionate to the commercial transaction under negotiation, and that to a large extent the loss may be mitigated by measures available to the data owner, such as backups and redundancy.

Contract drafters and negotiators should consider such exclusion terms carefully, in light of the value of their data, the risks and benefits of the contract under consideration, the technical redundancy and resilience measures in place for the data, and their own insurance arrangements.

Patrick Sefton
Brightline Lawyers
13 March 2018