FOREWORD: ABANDON ALL HOPE?

GRAHAM GREENLEAF AM*

The occurrence of 'communications surveillance' and 'big data' in a title brings privacy issues immediately to mind, but the articles in this Issue have a much broader scope than privacy, though it remains as one underlying thread. Issues of discrimination, of automated decision-making, of democracy, and of the public's right to access information, also thread through these pages. In one form or another, 'big data' (and not only in its fashionable form as 'Big Data') is a common element of all these articles. Big data is a puzzling concept, sometimes described as 'an all-encompassing term for any collection of datasets so large and complex that it becomes difficult to process using on-hand data management tools or traditional data processing applications.'[1] It is therefore not surprising that 'big data analytics' takes on an aura of mystery: tools to process that which cannot be processed. In the reality of particular company or agency operations, big data is often more accurately described as 'bigger data' than was previously attempted to be processed, drawn from more disparate sources (including transactional data, and social media data), with different structures or lack of them.[2] Nevertheless, reverence surrounds the conclusions or inferences generated by its mysterious 'analytics', the tools by which it analyses this previously unprocessable data.[3]

The place to start understanding big data is therefore the logic of its processing, which is the focus of two articles in this Issue. Bennett Moses and Chan, in 'Using Big Data for Legal and Law Enforcement Decisions: Testing the New Tools', examine two related domains of application, decision-making in the legal system and in law enforcement. They focus on 'data analytics', arguing that its techniques are often still rule-based, though utilising machine learning, and

* Professor of Law and Information Systems, UNSW Australia.

1 Wikipedia, *Big Data* (20 August 2014) <http://en.wikipedia.org/wiki/Big_data>. See Lyria Bennett Moses and Janet Chan, 'Using Big Data for Legal and Law Enforcement Decisions: Testing the New Tools' (2014) 37 *University of New South Wales Law Journal* 643, 650–2 for a discussion of definitions.

2 A more modest definition is '[b]ig data is a collection of data from traditional and digital sources inside and outside your company that represents a source for ongoing discovery and analysis.': Lisa Arthur 'What Is Big Data' on Forbes, *CMO Network* (15 August 2013) <http://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/>.

3 See boyd and Crawford's inclusion of 'mythology' in their definition of 'big data': danah boyd and Kate Crawford, 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon' (2012) 15 *Information, Communication & Society* 662, 663.

that the rules are usually opaque to end users. Their discussion of possible applications to legal decision-making makes clear that many of the issues that proponents of big data techniques will have to face have been well-trodden 30 or more years ago, and that the legal profession has been very discriminating in the tools it has adopted. In relation to police information systems, including sentencing databases, the result has also been to only accept those tools that augment human decision-making rather than replace it. Anticipating the next article to be discussed, they observe that the correlations brought out by data analytics may be unusable because of public policies concerning discrimination, but it is often not transparent that such correlations are being acted upon.

Bennett Moses and Chan refer to 'expert systems' and correctly stress that it is important to understand 'the continuities and differences between big data and precursor technologies'. This recalls the logic of a previous iteration of computing practices that were going to 'make everything different'. Japan's 'Fifth Generation' project of the early 1980s generated a similar amount of hype, and many believed it would make Japan dominant in computing within a decade. Parallel processors and logic programming techniques would enable inferences to be drawn from (newly possible) 'massive' databases. After expenditure of US$400 million (in 1980s dollars), it 'fizzled to a close' and resulted in only some minor changes to business practices – no revolution.[4] History does not usually repeat itself, but it gives repeated warnings about hubris.

The practices of data analytics in the employment context are given an illuminating analysis by Burdon and Harpur in 'Re-conceptualising Privacy and Discrimination in an Age of Talent Analytics'. The big data they write about is 'big' in the unprecedented intensity of its scrutiny of the actions of employees or potential employees, rather than 'big' in the sense of involving vast numbers of people or quantities of data. But it shares with other big data practices the advocacy of collection and analysis of all data that can be obtained – no matter how seemingly irrelevant – in the hope (and hype) that statistically-based analyses can produce non-intuitive correlations between aspects of employee/applicant behaviour (eg, web browser selection) and desired traits in the workplace (eg, punctuality). Burdon and Harpur analyse these practices against anti-discrimination laws in countries such as Australia, where certain 'protected attributes' (eg, race, sex, able-ness, age, genetic predisposition) cannot generally be the basis of decisions in particular relationships, including employment. Their argument that it is almost impossible for these laws to be applied when decisions are made on the basis of 'talent analytics' is very important, if we are to preserve the hard-won social policies represented by anti-

4 See Edward A Feigenbaum and Pamela McCorduck, *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World* (Addison Wesley Publishing, 1983); Andrew Pollack, '"Fifth Generation" Became Japan's Lost Generation', *The New York Times* (online), 5 June 1992 <http://www.nytimes.com/1992/06/05/business/fifth-generation-became-japan-s-lost-generation.html>. For an overview, see Wikipedia, *Fifth Generation Computer* (29 July 2014) <http://en.wikipedia.org/wiki/Fifth_generation_computer>.

discrimination laws. If the hidden heuristics of emerging employment practices start to mean that 'data is destiny' and it is usually almost impossible for either data users or data subjects to know even what data is being used to make decisions, this also points to the types of problems that 'big data analytics' are likely to bring in other contexts (such as those discussed by Bennett Moses and Chan).

An aspect not pursued by Burdon and Harpur is whether the intensive personal data collection they describe as the starting point of talent analytics, the collection of 'a log of the employee's activity throughout the working day', should be acceptable. To what extent should constant workplace surveillance be allowed, even if it is *only* of the metadata comprised in the continuous details of an employee's use of company systems (whether computers or toilets) in an industrial relations system? To what extent does or should it breach the 'data minimisation' principles of data privacy laws?

Big data takes on another meaning in the context of government surveillance powers, and two articles in the Issue focus on those powers. In 'Surveillance, Big Data and Democracy: Lessons for Australia from the US and UK', de Zwart, Humphreys and van Dissel doubt that current concepts of privacy based on controlling collection, storage, and use of categories of data, are now capable of controlling the types of practices involved in 'big data analytics'. They give a detailed account of recent case law in the USA (mainly concerning government surveillance practices, resulting in few controls) and in the UK (where cases leave unresolved the extent to which data protection laws can restrain search engines). The 'woefully underdeveloped' state of Australian privacy law, in all its forms, is then sketched.

Since July 2014, there has been extensive debate occurring in Australia of the Abbott government's active pursuit of data retention legislation, as part of the more extensive legislative review of data surveillance and telecommunications interception powers. 'Metadata' has decisively entered the Australian lexicon through the contradictions between various ministers and security officials concerning what the term means. Lachmayer and Witzleb's article, 'The Challenge to Privacy from Ever Increasing State Surveillance: A Comparative Perspective', gives a broad international background to this debate, comparing equivalent developments and debates in Europe and the USA with what is occurring in Australia. The dismal lack of constitutional or statutory protections that they describe underlines the need for new protections to be included in the current Australian legislative reforms – but there is no indication that they are likely to occur.

Lachmayer and Witzleb regard the *Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data* ('*Convention 108*')[5] as the most likely candidate to grow into an international agreement on

5 Opened for signature 28 January 1981, CETS No 108 (entered into force 1 October 1985).

data privacy, an assessment with which I agree.[6] However, they may be unduly pessimistic in considering that the data privacy standards developed by the European Union – particularly when they are strengthened by the eventual adoption of a data protection regulation – will not gain international traction. The history of 40 years development of data privacy laws over what is now 105 countries shows the strong and growing adoption of the 'European' standards shared by the *EU Directive*[7] and *Convention 108*,[8] to which I will return in concluding.

Chelsea (Bradley) Manning, Julian Assange and Edward Snowden raise quite different big data issues, where very big sets of government-held data are made available to the public. The legal consequences in Australia of this enabling of sousveillance of government actions are meticulously analysed by Hardy and Williams in 'Terrorist, Traitor or Whistleblower? – Offences and Protections in Australia for Disclosing National Security Information'. They find few protections in Australian law, at least where information related to national security is concerned, irrespective of the public interest justifications offered for disclosures. Big data only seems to be allowed to flow in one direction.

A common element in all the articles in this Issue is pessimism: little enthusiasm for the promises of big data, and many concerns about its dangers; and shared dismay at the inadequacy of privacy laws to deal with the problems raised by it, or by surveillance practices. This is not surprising. It is a common perception in 2014, a year after the Snowden revelations started, that individual privacy is more under threat that ever before. Media reports, and most people they interview, can reel off lists of concerns: search engine tracking, social media, cloud computing, state surveillance, data spills, and so on – even 'big data'. Are all of these things simply developments on a continuum of gradually intensifying and expanding surveillance? Has it just been occurring incrementally since the late 1960s, when concerns were first raised about 'databanks' of personal information?[9] Or has there been some fundamental discontinuity in surveillance or social practices in recent years, so that we are now living in a society where threats to privacy have passed some 'tipping point' and its protection is now impossible? If so, then what change or changes are the root cause of the discontinuity, and when did they occur?

---

6 Graham Greenleaf, 'A World Data Privacy Treaty? "Globalisation" and "Modernisation" of Council of Europe Convention 108' in Normann Witzleb et al (eds), *Emerging Challenges in Privacy Law: Comparative Perspectives* (Cambridge University Press, 2014) 93.

7 *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data* [1995] OJ L 281/31.

8 Graham Greenleaf, 'Sheherezade and the 101 Data Privacy Laws: Origins, Significance and Global Trajectories' (2014) forthcoming *Journal of Law, Information and Science* <http://www.jlisjournal.org/abstracts/greenleaf.23.1.html>.

9 For early important analysis, see especially James B Rule, *Private Lives and Public Surveillance* (Allan Lane, 1973). See also Alan F Westin and Michael A Baker, *Databanks in a Free Society: Computers, Record-Keeping and Society* (Quadrangle Books, 1972).

A brief and incomplete catalogue of factors that have made the preservation of privacy far more difficult demonstrates both the current severity of the problems and the difficulty of identifying whether some changes have made a qualitative difference.

(i) *Environmental factors* which make the technical realisation of other threats more easily achievable. These include the progressive elimination of processing costs (Moore's law); the progressive elimination of data storage costs; and the explosion of online transactions.

(ii) *Collection and generation of usable personal data.* Developments include digitisation of new/more personal data (biometrics, photos, old texts); visual data collection (and use/misuse of data collected); voluntary disclosure (social media, photos, user-generated content generally, 'gamification' of interactions); big data aggregations; personalised interaction without identification (intrusions, mobile devices, geolocation techniques); de-anonymisation of transactions (both sectors, IDs for everything, removal of anonymous options); re-identification techniques (destruction of previous anonymity); government IDs (increased scope even if there is no universal ID); and commercial personal data collection (cookies, etc).

(iii) *Intensified processing and use of personal data* – including by more powerful processing of personal data (analytics, search and ranking, etc); through commercial interconnection (eg, advertising syndication, internet of things); and by state interconnection (data matching, ID systems).

(iv) *Increased retention and reduced security of personal data* – including by permanent retention of personal data (both on our own devices, and on third-party devices); by the endemic failure of security systems (increased risks through internet connectivity and the unknown problems of legacy systems); and the risks of massive systems failure and data unavailability.

(v) *Increased disclosure and transfer of personal data* through state surveillance (Snowden revelations); foreign state surveillance (Snowden again); international data mobility (cloud services, etc); malicious hacking (markets in stolen personal data, botnets); and unintended data breaches.

Of course, any one of these factors can make many of the other factors more damaging. But which are qualitatively different from what was there before? If we look back to around 2001, when the first 'internet bubble' burst, one notable difference from the post-millenium Internet 2.0 is that in Internet 1.0 the user was not 'the product', and it was not based on behavioural marketing. In that first internet boom of 1994–2000, there were obsessions that now seem as quaint as tulipmania, such as inflated value of domain names, and the related attraction of particular 'portals'. It was still believed that particular classes of users would be attracted to particular websites, at which point marketing to them could take

place. Since then, internet commerce has been to some extent redesigned around the maxim 'when the service is free, the user is the product,'[10] and the aggregation of information about individuals for the purpose of onselling that information for other marketing activities has come to predominate. Simultaneous with this has been the explosion of both 'voluntary' disclosure of personal data – user-generated content, social media, etc – and of mobile devices (and geolocation) which has made constant participation possible despite (and to include) physical location. Julie Cohen identifies that this 'participatory turn' increasingly involves 'gamification' of user interactions, staged competition with known and unknown peers in order to increase the motivation to disclose personal data.[11] This is part, she argues, of a developing ideology that positions surveillance as the partner of innovation. No longer a potential danger to be regulated, it is positioned as a source of innovation to be encouraged – a prospecting tool of a new biosphere of personal data. On this hypothesis (and it is no more than that) 'big data analytics' are then a key part of the seismic shift, but not the shift itself. Coincidental in timing with the start of any millennial change were the 2001 attacks on New York City, and the shift to vastly intensified state surveillance, discussed in this Issue. The interconnections between these major changes in both private sector and public sector surveillance, both ideological and technical, will take a great deal of unravelling.

To conclude, let me return briefly to pessimism about privacy, and data privacy laws in particular. Across the globe, the content of data privacy laws and the means of enforcing them, have been emerging for little over 40 years, through the interaction of iteratively stronger developments of international standards and national laws. Neither strand is complete and both continue to strengthen. Many aspects of the proposed EU Regulation,[12] and likely parallels in *Convention 108*, will be inimical to privacy-invasive business practices, including those based outside Europe. These include more explicit data minimisation, the 'right to be forgotten', data portability, 'privacy by default', stronger extraterritoriality, local representative requirements, and fines proportional to business size. Some similar reforms are already being introduced outside Europe, in countries like South Korea. Such European changes, if the past is a guide, will gradually diffuse worldwide.[13]

Nevertheless, strengthening privacy laws are everywhere overshadowed by the ability (legal or not) of US-based companies to 'hoover up' the personal data

10 See, eg, Daniel Newman 'There Is No Privacy on the Internet of Things' on Forbes, *CMO Network* (20 August 2014) <http://www.forbes.com/sites/danielnewman/2014/08/20/there-is-no-privacy-on-the-internet-of-things/>.

11 Julie E Cohen, 'The Surveillance-Innovation Complex: The Irony of the Participatory Turn' in Darin Barney et al (eds), *The Participatory Condition* (University of Minnesota Press, 2015) forthcoming.

12 See, eg, *Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)* [2012] COM(2012) 11.

13 Graham Greenleaf, 'The Influence of European Data Privacy Standards outside Europe: Implications for Globalisation of Convention 108' (2012) 2 *International Data Privacy Law* 68.

of people in the rest of the world, and to process and use it with few restrictions. At present, privacy standards in other countries do not matter much if personal data can be liberated to the US 'Safe Harbor'. It should not be forgotten that a similar situation prevailed a century ago when the US was the pirates' harbour of the copyright world, to the despair of authors and countries with then-emerging 'international standard' copyright laws. National attitudes to laws can change 180 degrees with changes in business models, as occurred with the US and copyright. It finally joined the *Berne Convention*[14] after 102 years in 1988. The prevailing US model of an internet where 'the user is the product' is not necessarily permanent. However, to stop it becoming so, it will either take a second internet bubble to burst, or a concerted effort by the rest of the world to reject privacy-invasive business practices. Neither is impossible, nor likely to occur rapidly.

14 *Berne Convention for the Protection of Literary and Artistic Works*, opened for signature 9 September 1886, [1901] ATS 126 (entered into force 5 December 1887), as last revised by the *Paris Act Relating to the Berne Convention for the Protection of Literary and Artistic Works*, signed 24 July 1971, 1161 UNTS 30 (entered into force 15 December 1972), and further amended by the *Amendments to Articles 22 and 23 of the Paris Act of 14 July 1971 of the Berne Convention for the Protection of Literary and Artistic Works*, signed 28 September 1979, [1984] ATS 40 (entered into force 19 November 1984).