Privacy in the era of Big Data:
Meaning and implications for Aotearoa New Zealand

by Elise Antoine[1]

Last year the Facebook-Cambridge Analytica scandal gripped many nations, as personal data from millions of Facebook profiles (whose 64,000 New Zealanders) have been harvested by the consulting firm for political campaigning purposes and used without consent. It has now become evident that the rise of big data put the fundamental right to privacy at risk.

As aggregated data can provide insights into the life of groups and individuals, new platforms for collecting and storing data are brought to better inform policy-makers. An example of this is the Integrated Data Infrastructure (IDI), which is New Zealand's largest database about people and households. The analysis of data collected by governments for policy making is a common practice. However, what is different in the case of the IDI is the volume and diversity of data collected from and about people, which takes us to a significant issue in big data: privacy.

The notion of privacy is complex. Simply put, it is about being free from any intrusion in your personal life. This freedom involves many elements, but in the digital context, it particularly implicates the capacity to control information about oneself and the possibility of being not identifiable, *i.e.* anonymous.

The revelations by Edward Snowden in 2013 have rendered surveillance practices more visible than ever, but privacy is not only about the content of communication in social networks or the information contained in mobile smartphones. Privacy is also involved in surveys and administrative data[2]. Indeed, when aggregated and analysed, these types of data can provide insights into the behaviour of individuals and groups. The creation of massive datasets has, therefore, far-reaching ethical implications.

The ethical implications include the potential for privacy breach. The IDI has vulnerabilities when it comes to guaranteeing the privacy of people whose data collect, particularly because anonymity may be broken. This point lies at the heart of this paper.

---

[1] Elise Antoine, who comes from France, is an intern at UNA NZ. She is a recent graduate from Panthéon Sorbonne University, where she studied Political Science. Elise has developed an interest in digital technologies throughout her academic background and her internship at UNA NZ. She plans to strengthen her expertise in this area in the future.

[2] Defined as data collected by government departments for the purposes of registration and record keeping (*e.g.* income tax).

To understand how privacy can be at risk in the IDI, we further elaborate on the dimensions of privacy and how they interact (*i.e.* control of personal data and anonymity). We also need to know how privacy can be better protected. This paper does not intend to address all the ethical issues raised b*y* Big Data, but to provide a critical review of data privacy issues in New Zealand, specifically in the context of the IDI.

Based on the analysis of official documents primarily, this paper first examines how privacy is protected in New Zealand. Then it looks at the European Union data privacy legislation, which provides the highest protections so far. In the second part, the paper discusses the IDI with an emphasis on its safeguards and weaknesses and explores further areas of research to conclude with the main findings of the critical review.

I. Privacy and Big Data: definitions, issues and regulations

A. Big data, Privacy and the United Nations

The United Nations have adopted 16 Sustainable Development Goals (SDGs)[3] to achieve a more sustainable future. As big data help to inform policy-makers, it can contribute to each of these Goals. Nevertheless, human rights have to be protected to realize the opportunities that big data presents.

The right to privacy has been upheld as a fundamental human right in the Universal Declaration of Human Rights (UDHR) and the International Covenant on Civil and Political Rights (ICCPR).

« No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks ». (The United Nations, 1948, art. 12).

The problem societies are facing today is that digital technologies are integrated into every sphere of our lives, and consequently, the space for being free from any « interference » is shrinking. More specifically, big data enable the collection and analysis of a massive amount of personal information[4], very often without the consent of concerned individuals. Research and policy-making purposes legitimate the collection and storage of personal information supposing that individuals have their identity protected. However, because a lot of information is becoming available, it becomes more difficult to remain anonymous.

---

[3] Exemples of these Goals: ensure healthy lives and promote well-being for all at all ages, reduce inequality https://www.un.org/sustainabledevelopment/sustainable-development-goals/

[4] Personal information is defined as information which allows identifying an individual, like the driver's licence number.

These issues are *global*: they can affect anyone, in any country. Global issues require global responses, *i.e* solutions that go beyond national boundaries. The United Nations General Assembly adopted its first resolution on the right to privacy in the digital age in 2013, affirming that «the same rights that people have offline must also be protected online» (United Nations General Assembly, 2013). However, there is currently no framework that regulates digital technologies and protects privacy on a global scale.

We will now focus on how privacy is safeguarded in two regions: New Zealand and the European Union.

B. The right to privacy in Aotearoa New Zealand

The rise of big data requires protection of personal information. In New Zealand, personal information is protected through the Privacy Act of 1993.[5] The Act contains 12 information principles that aim to prevent data breach (also called privacy breach). A data breach is a loss or unauthorised use of personal data. The disclosure of personal information constitutes a data breach too. These breaches can result in a financial loss or emotional distress, for example, for patients whose diagnosis have been publicly exposed. A privacy breach jeopardises human dignity.

Notably, the Act seeks to protect individual privacy, and as such, is based on the capacity of the individuals to manage their data (*e.g.* right to access information and correct it). However, there is no specific framework when it comes to data from and about Māori. The Act recognises individual privacy but not the collective one.

The number of privacy breaches, particularly in terms of unauthorised use data, has increased over the last few years. The Facebook-Cambridge Analytica scandal was an example. More recently, the Privacy Commissioner has revealed that the Ministry of Social Development was collecting personal data of beneficiaries including text messages, police and banking records.[6]

Given the increase of privacy breaches, last year a new Privacy Bill was introduced into Parliament to replace the Privacy Act. When a privacy breach has occurred, the individual affected is currently responsible for making a complaint to the Privacy Commissioner. The new Privacy Bill shifts the responsibility. The agency that collects the data will now have to notify both the individual and the Privacy Commissioner when a breach that caused harm (or risk of harm) has happened.[7] This change is essential to increase transparency and accountability in data use.

---

[5] Here we need to note that there is no general right to privacy in the New Zealand legislation. The right to privacy is protected in Privacy Act (which only deals with the personal information) and the Fencing Act 1978 (which emphasizes the right to enjoy a private space)

[6] https://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=12231319

[7] Privacy Bill 34-2 (2018), Part 6

C. Europe and the General Data Protection Regulation (GDPR): why is it important?

In the last few years, privacy laws all around the world have been reformed. The most influential of those reforms has been the GDPR, which came into force in the EU countries in May 2018. The GDPR sets the rules applicable to the processing of EU residents personal data, which apply both to private and public sectors. Although the GDPR has affected many countries outside Europe, it is not sufficient to impose worldwide rules on privacy. [8]

The NZ Privacy Act reform will bring the country close to European regulation, as it acknowledges data subjects' rights and obligates to report privacy breaches. However, the NZ Privacy Act falls behind the GDPR on significant issues. The first one is that the GDPR defines personal data in a broader sense: it refers to any data that can be linked to a person, even if only in combination with other data (*e.g.* location data from a mobile phone). Additionally, with article 17, the GDRP recognises a « right to be forgotten » that extends individuals' capacity to control their data. Indeed, individuals can ask for the erasing of their data, which at the same time, strengthens their capacity to withdraw consent. Also, article 89 of the regulation allows the use of personal data for research purposes when safeguards preventing re-identification of individuals are implemented.

These national and international frameworks delineate the privacy issues brought about by big data and underpinned in the IDI. We will now focus on this database by analysing how it operates and what issues it raises. It is important to engage a reflection on the IDI given the personal and sensitive information it contains about people, and consequently, the potential for misuse and abuse of such information.[9]

II. The Integrated Data Infrastructure

A. What is the IDI, and how does it work?

The IDI is a large research database from and about people. It has data on different topics such as education (*e.g.* students enrolled at primary school) health (*e.g.* cancer registrations) and justice (e.g. recorded crimes) among many others. The data comes from government agencies, Statistics NZ surveys, and non-government organisations.

---

8 Many features can explain this influence. Firstly, Any agency, which deals with data from EU citizens, have to comply with the GDPR (known as extra-territorial effect of the GDPR). Secondly, data from EU countries can only flow with third countries that have an adequate level of data protection. Finally, the EU is often described as a « normative power », given its capacity to export rules. How the GDPR has affected countries outside Europe would merit further discussion.

9 See how the Chinese government is using big data to monitor and evaluate every activity of citizens. http://time.com/collection/davos-2019/5502592/china-social-credit-score/

The idea behind the creation of the IDI was to look for relationships between different dimensions, such as health outcomes and economic revenue, for example. Thus, the data in the IDI is stored for other purposes than those for which it was initially collected. It is believed that the discovery of such relationships might improve public policies.

Once data are collected from different agencies, the statisticians match the data in one source to the data they believe are the most likely associated with that person in another source (process of linkage). Then, data are de-identified, which means information like names and addresses are removed; hence, information becomes anonymous.

B. Safeguards and weaknesses

The IDI operates under a clear purpose: improving the quality of public services by enabling research based on linked data. Researchers (from government departments or universities) must demonstrate how their project contributes to this purpose. However, there is no independent ethics committee for reviewing the projects, and only Statistics NZ is responsible for accepting or refusing the proposals.

Importantly, privacy is considered at different stages. Potential risks to individual privacy are considered before adding data in the infrastructure. Also, researchers can only access data in a secure environment without internet or USB access after attending privacy and confidentiality training.[10] Finally, data is de-identified, so individuals cannot be recognised. The IDI is an example of how researchers and policy-makers can use aggregated data ethically.

Nevertheless, New Zealanders could be further involved to decide how using data, particularly Māori who recognise collective privacy. Above all, anonymity is potentially at risk. The IDI is based on *big data* and, researchers may recognize an individual due to the fact that they know a large number of different features about the person.[11] The more data is aggregated, the more individuals can be identified. New Zealand is a country of just over 4.5 million inhabitants, which facilitates the re-identification even more. There is currently a gap in the New Zealand legislation, as in most privacy laws, since re-identification is not taken into account.[12]

---

[10] Among others, they learn how to use statistical methods to protect the confidentiality of information.

[11] For a more detailed case study, see Barth-Jones, Daniel, *The 'Re-Identification' of Governor William Weld's Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now* (July 2012). Available at SSRN: https://ssrn.com/abstract=2076397 or http://dx.doi.org/10.2139/ssrn.2076397

[12] Even under the GDPR, it is unclear what safeguards prevent re-identification and what are the legal consequences of the intentional re-identification.

III. Big data is massive; so are the ethical questions

Because individuals are not identifiable does not mean that harm cannot occur. Groups of people (*e.g.* social, ethnic or religious groups) can be identified and flagged, which may lead to discriminatory practices. Further research could, therefore, focus on the challenges of an « IDI-based policy »: how data is interpreted and used in policy-making? What are the implications?

The IDI operates under secure principles, but there is still room for improvement in data governance and above all, to guarantee anonymity. Although the new Privacy Bill does not currently address the re-identification issue, it significantly improves the transparency and accountability in data use. New Zealand, like the EU, seeks to counter data breaches and protect the right to privacy « offline ». These objectives remain challenging as every type of data when aggregated and analysed might threaten our right to control and limit access to our personal information, as well as our right to enjoy anonymity.

Privacy is essential for a person to be themselves, *i.e* to develop develop unique individuality. As a fundamental human right, privacy requires constant debates and efforts to keep it safe.

References:

Andrade N. (2010). Data Protection, Privacy and Identity: Distinguishing Concepts and Articulating Rights. *6th International Summer Schoo*l (ISS), Helsingborg, Sweden. pp.90-107. 10.1007/978-3-642-20769-3_8 . hal-01559453

General Assembly, resolution 68/167 The right to privacy in the digital age. (18 December 2013). Available from undocs.org/A/RES/68/16

McDermott, Y. (2017). Conceptualising the right to data protection in an era of Big Data. *Big Data & Society*.

New Zealand Human Rights Commission. Privacy, Data and Technology: Human Rights Challenges in the Digital Age. (2018).

Privacy Act 1993, No.28. Retrieved from http://tekete.ara.ac.nz/file/9a02a036-ab19-4369-8d09-8b2894708014/1/APA%206th%20Ed%202014%20Final.pdf

Privacy Bill 2018, N.34-2. Retrieved from http://www.legislation.govt.nz/bill/government/2018/0034/latest/whole.html

Privacy Commissioner's Submission on the Privacy Bill to the Justice and Electoral Select Committee. (2018).

Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). (2016) Official Journal, L 119.

Terwangne, Cécile, de. (2014). *The Ethics of Memory in a Digital Age.* Palgrave Macmillan UK.https://ec.europa.eu/jrc/en/publication/right-be-forgotten-and-informational-autonomy-digital-environment

The United Nations General Assembly. (1966). International Covenant on Civil and Political Rights. *Treaty Series*, 999, 171.

The United Nations General Assembly. (1948). Universal declaration of human rights (217 [III] A). Paris.

Vayena, E. & Tasioulas, J. (2016). The dynamics of big data and human rights: The case of scientific research. *Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences*. 374. 20160129. 10.1098/rsta.2016.0129.