Is the current definition of personal information enough to protect individuals from privacy harms?

Marcin Betkier, Natasha Mazey, Reuel Baptista

Privacy Foundation New Zealand - Privacy in the Internet Economy Working Group

22 March 2021

This paper considers and furthers the debate provided in the thought-provoking paper *Individuation: Re-imagining Data Privacy Laws to Protect Against Privacy Harms* by Anna Johnston.[1] Johnston identifies a potential issue with existing data privacy (and data protection) laws in countries that use the concept of identifiability of a person as a criterion to classify information (or data) as 'personal' in order to protect it. Simply put, does protecting identifiable personal information protect individuals from privacy harms? Johnston argues it does not; identifiability is not the right criterion because it does not address scenarios where individuals can suffer privacy harms even though their identity is not known or knowable. Those harms "can also arise from *individuation*: the ability to disambiguate or 'single out' a person in the crowd, such that they could, at an individual level be tracked, profiled, targeted, contacted, or subject to a decision or action which impacts upon them".[2]

Johnston is undoubtedly right that the definition of personal information is crucial. However, this definition and the threshold at which data protection applies is not defined strictly. Instead, the answer is influenced by context and circumstances. In principle, we agree with Johnston's argument about the necessity of protecting individuation. However, we explore and argue that existing legal protections and practises provide a reasonable level of protection without altering the definition of personal data.

Policy background

Policy considerations around personal information is the most important point to start with. Privacy and data protection statutes balance competing values (or goals). On one side there is a need to protect individuals' right to privacy. On the other side is the need for 'free movement of personal data' (the General Data Protection Regulation, GDPR), also known as the interest of data holders (Australian Privacy Act)[3] or the promotion of electronic commerce (Canadian PIPEDA).[4] International

[1] Anna Johnston "Individuation: Re-imagining Data Privacy Laws to Protect Against Digital Harms" (2020) 6 Brussels Privacy Hub Working Paper.
[2] Above, at 1.
[3] Privacy Act 1988 (Cth).
[4] Personal Information Protection and Electronic Documents Act (PIPEDA).

instruments like the OECD Privacy Guidelines[5] and the Council of Europe Convention 108[6] point to 'the free flow of information'. Also, there is a wider range of policy considerations to incorporate into this balancing exercise. There are important social interests in preserving privacy, for example, Cambridge Analytica's use of extensive data harvesting to target political advertising undermined not only citizens' privacy rights, but citizens' confidence in the democratic process and the democratic process itself.

So, protection of the individual is not the only policy goal. Privacy statutes, particularly the scope of the definition of personal information, need to reflect that. The law has to protect the privacy of personal information, maintain the workability of the regulation and balance values such as the societal benefits of sharing the information and economic interests. This could be a practical challenge, considering the potential scope of the definition of personal information covers '(any) information about (relating to) an identifiable person (individual)'. If the law sets the scope too wide, it may start to be a 'law of everything',[7] because of the broadness of the notion of 'identifiable', and also broadness of 'any information' and 'about' (or 'relating to'). That could have huge, and potentially disproportionate, consequences. For example, sets of data for everyday use could be turned into datasets of personal information with limited availability for use. That could also mean practical unworkability of privacy and data protection concepts and, paradoxically, less protection. Conversely, if the scope is too narrow, it could mean that individuals do not have the necessary level of protection. In light of these challenges, the definition of personal information needs to be interpreted in a way which sets the scope of that definition 'just right'.

Note that the scope of the definition needs to be continuously reviewed to account for developments in technology and data practices. For example, organisations have acquired tools and learned new methods and tactics to merge, analyse and profile data. Consequently, data which would not have been able to identify individual a few years ago, now may be perceived as identifiable personal information.

The convergence of identifiability

The most important criticism of Johnston's paper seems to be that the current law generally does not necessarily rely on 'the individual being findable and identifiable in a legal sense'.[8] She seems to suggest that identifiability boils down to capturing a person's name and other credentials relevant to match and identify a person to an individual's legal or national identity.[9] However, a person's name and/or surname is not the only way of identifying him or her.[10] It may be enough to know other unique identifier(s), for example, a state-assigned identity number or a pseudonym. Alternatively, there may be a collection of characteristics that makes a person identifiable by people who know or have additional information about him or her (e.g. physical appearance, job role, lifestyle).[11] This

[5] OECD *Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data* (2013).
[6] Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data, Council of Europe 108 European Treaty Series (adopted 17–18 ,May 2018).
[7] E.g. Nadezhda Purtova "The law of everything Broad concept of personal data and future of EU data protection law" (2018) 10 Law, Innovation and Technology 40.
[8] Johnston, above n 1, at 9, 14.
[9] See, for example, definition of legal identity in United Nations "UN Legal Identity Agenda" <https://unstats.un.org>, "Legal identity is defined as the basic characteristics of an individual's identity. e.g. name, sex, place and date of birth conferred through registration and the issuance of a certificate by an authorized civil registration authority".
[10] Also, especially for popular names and surnames, it may not uniquely identify a person.
[11] E.g. discussion in Katrine Evans "Personal Information in New Zealand: Between a Rock and a Hard Place" [2006] 8 at 2–4.

may be sufficient to identify, or re-identify, a person in a dataset.[12] The concept of individuation to address weaknesses in the understanding of identifiability, as associated with legal or national identity, could be appropriate. However, we note that many definitions of personal information (see for example, the GDPR or New Zealand Privacy Act 2020)[13] do not require that this be linked to any legal or national identifier.

Identifiability is widely understood in the field of data and data privacy as the ability to sufficiently identify a single subject within a group of subjects.[14] For example, in relational databases, a 'key' is an attribute or *a group of attributes* (characteristics described in data) that enables the identification of a single entity or record in a table.[15] If records describe different people, identification could occur by finding the key, which consists of set of those attributes. This is also the understanding of the concept of identifiability by the European Data Protection Authorities described in their Opinion on the concept of personal information:[16]

> *a natural person can be considered as "identified" when, within a group of persons, he or she is "distinguished" from all other members of the group. Accordingly, the natural person is "identifiable" when, although the person has not been identified yet, it is possible to do it…*

This does not necessarily end with finding out a person's name.[17] Rather, the ability to distinguish a person from the members of the group is enough.

Therefore, the 'identity' of a person, besides its philosophical meaning, is a subset of attributes that allow a person to be identified in a set.[18] Also, the consequence is that one person may have many identities. That also seems to be congruent with the commonly used meaning of that word. For example, one of the authors could be identified by name and surname and place where he lives i.e. Marcin, Betkier, Wellington (domicile is not needed if we focus on New Zealand and Australia), but also using his ethnicity (Polish) and workplace (VUW Law School), maybe by the set of his particular hobbies or interests, and probably by the set of apps installed on his smartphone (as being specific for a person having two nationalities and that set of interests). The consequence is that there are many identities of the same person and that person can be identifiable in many different ways than by his or her name. Does it mean that we do not need 'individuation'?

Is 'individuation' needed for protection against 'singling out'?

We do not need the concept of individuation to protect individuals from harms resulting from being 'singled out'. This is because singling out, as addressed by individuation, should already be

---

[12] See more about the failures of anonymising the person through "de-identification" in Marcin Betkier and Natasha Mazey "The Ignorance of Anonymisation to Protect Privacy" (26 September 2020) Privacy Foundation New Zealand <www.privacyfoundation.nz>. A well-known example of this can be found in the New York Times; "A Face Is Exposed for AOL Searcher No. 4417749", 9 August 2006.

[13] That may be different under Australian statute which refers to a 'reasonably identifiable' individual, see Privacy Act 1988 (Australia), s 6.

[14] E.g. Andreas Pfitzmann and Marit Hansen "A Terminology for Talking about Privacy by Data Minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management" [2010] Technische Universitat Dresden at 30.

[15] E.g. "Key in a Relational Database" Network Encyclopedia <https://networkencyclopedia.com>.

[16] Article 29 Working Party *Opinion 4/2007 on the Concept of Personal Data* (WP 136 2007) at 12.

[17] For a legal analysis see above, at 14.

[18] Pfitzmann and Hansen, above n 14, at 30.

understood as identifiability. Johnston's paper agrees with this in relation to the GDPR,[19] but later returns to the concept that 'each of these elements [arguments for broader meaning of identifiability] still comes back to the idea of the person ultimately being identifiable in a *legal sense*'.[20] Her argument is supported by the reference to the *Breyer* case. In that case, the court found that even if state authorities knew only a dynamic IP address of the plaintiff's computer, he was identifiable because the information about his name and address was possible to obtain by them from the third party (Internet Service Provider).[21] Similarly, in *Planet49* (not cited by Johnston, but potentially supporting her argument) the court found online cookies to be identifiable personal data under the law because each cookie included an identifier which could be associated with the user's name and address that was collected when they registered online.[22]

Having said that, even if the idea of identifying a person by the means of his or her 'legal identity' was used by the court in *Breyer* and *Planet49*, courts do not *have to* use it to determine whether individuals are or were individually identifiable. Those data were simply available in those cases, but they are not always necessary. That understanding can be found in other court cases and in practice of Data Protection Authorities. For example, the same European Court of Justice in the famous *Lindquist* case in 2003 stated:[23]

> *…referring, on an internet page, to various persons and identifying them by name* ***or by other means, for instance by giving their telephone number or information regarding their working conditions and hobbies,*** *constitutes the processing of personal data…*

Similarly with the use of cookies, the UK Information Commissioner's Office recently stated that cookies are personal data '(…) where identifiers are used or combined to create profiles of individuals, **even when those individuals are unnamed**.'[24] Further, Google has been very recently fined by the French data protection authority for placing cookies without user's consent.[25] In New Zealand, the Human Rights Review Tribunal recognised that even when the individual is not named, there may be, depending on the context, a sufficient connection to that individual to justify a conclusion that the information is personal information 'about' that person.[26] Very similar findings were presented by the New Zealand Privacy Commissioner.[27] So, the wider understanding of 'identifiability' already exists and is used in practice.

Does the current definition of personal information cover the possibility of harming a group or a class of people?

Johnston's paper still makes a relevant point that data protection and privacy laws do not cover the possibility of harming a group or a class of people. This is because it is possible to harm a group of people without identifying (or individuating) them. In this sense, Johnston's examples of targeting

---

[19] Johnston, above n 1, at 13–14.
[20] At 14.
[21] *Patrick Breyer v Bundesrepublik Deutschland,* European Court of Justice, C-582/14.
[22] *Planet49,* European Court of Justice, C-673/17 at [45].
[23] *Bodil Lindquist,* European Court of Justice, C-101/01 at [27].
[24] Information Commissioner's Office *Guidance on the use of cookies and similar technologies* (2019) at 19.
[25] "Cookies: financial penalties of 60 million euros against the company GOOGLE LLC and of 40 million euros against the company GOOGLE IRELAND LIMITED | CNIL" <www.cnil.fr>.
[26] Director of Human Rights Proceedings v Hamilton [2012] NZHRRT 24 at [41],
[27] See paragraphs [10]-[14] in Privacy Commissioner *Advisory Opinion Whether addresses of fire incidents are personal information* (AO 001/2016 2017) at 3–4.

people that enter abortion clinics,[28] or distinguishing people on the basis of their 'racial affinity' or 'psychological vulnerability'[29] are on point. But, both data privacy and European data protection laws are intrinsically connected to the concept of harm to the *individual* and are not well equipped to address harms to the *group* that are not linked with identifiable individual.

There is literature related to 'group privacy' which aims to describe the issues that arise around personal information relating to a group of people.[30] For example, there are group privacy problems relating to DNA data that describe a family rather than more than just one person. Also, indigenous concepts of privacy often relate to information relating to a group of people (a tribe or larger family).[31] However, those problems, unfortunately, cannot be easily solved by the change of definition proposed by Johnston. This is because they are not under the concept of 'singling out' regardless of the fact whether it is included in the current definition of personal information or not. They need a deeper research; perhaps an overhaul of existing data privacy/protection laws or developing existing laws (anti-discrimination laws, for instance) to help address group privacy issues.

Summary

Johnston's paper identified two key problems: the harm arising from 'individuation' which, as the original intent of the privacy legislation would suggest, should be covered, and the harm arising because of targeting a group or class of persons. We argue that the first issue may not need to be resolved as suggested by Johnston. This is because the definitions of personal information used by statutes are functional, technology-neutral, and capable of addressing the ability to 'single out' a person *within* the concept of identifiability. However, the second problem, which considers the risk of harm to groups or a class of people that can be identified based on a group of attributes, may require much deeper discussion and changes to privacy and data protection legislation (and possibly anti-discrimination laws) to address this issue.

About the Privacy Foundation New Zealand

The Privacy Foundation New Zealand was established in 2016 to protect New Zealanders' privacy rights, by means of research, awareness, education, the highlighting of privacy risks in all forms of technology and practices, and through campaigning for appropriate laws and regulations. Its membership has a diverse range of professional, academic and consumer backgrounds and the Foundation regularly lends its collective expertise to comment on proposed regulation or programmes in the media or by participating in consultation processes.

[28] Johnston, above n 1, at 7.
[29] At 11.
[30] E.g. Linnet Taylor, Luciano Floridi and Bart van der Sloot *Group Privacy* (Springer International Publishing, Cham, 2017); Edward J Bloustein *Individual & Group Privacy* (2nd ed, Transaction Publishers, New Brunswick, 2003).
[31] Consider, for example, whakapapa (the line of descent from ancestors) for Māori.