CANON Canadian Anonymization Network

The Hon. Lisa M. Thompson
Minister, Ministry of Government and Consumer Services
College Park, 5th Floor, 777 Bay Street
Toronto, Ontario M7A 2J3

Sent via email to access.privacy@ontario.ca

October 16, 2020

Dear Minister Thompson,

Thank you for the opportunity to provide comment on the Ministry of Government and Consumer Services (MGCS)'s public consultation, *Reforming Privacy in Ontario's Private Sector* and associated *Privacy Reform Discussion Paper* (the "Discussion Paper").

We are writing on behalf of CANON, the Canadian Anonymization Network. Launched in May 2019, CANON is a not-for-profit organization whose members are comprised of large data custodians from across the public, private, and health sectors. The Network's primary purpose is to promote the use of privacy-enhancing technologies (including de-identification) in Canada as privacy-respectful means of supporting innovation and leveraging data for socially and economically beneficial purposes.

This submission is not intended to state a formal policy position on behalf of CANON or its members in respect of the entire Discussion Paper. Rather, the present submission is limited to comments about de-identification[1] issues more specifically.

The members of CANON are encouraged to see that the MGCS has included a recognition of the potential benefits of de-identification in its Discussion Paper, including that it can be an enabler of privacy-protective data-sharing for innovation. In brief, we would like to voice our mutual interest in continuing to advance the policy debate on the important virtues of de-identification and other privacy-enhancing technologies (such as synthetic data) as a privacy-protective measure and offer some suggestions and points of clarity. To that end, we would be pleased to make ourselves available for any follow-up questions or discussion of the points raised below, or any other matter for which the collective experience and/or expertise of CANON members may prove to be helpful.

We have divided our submission into four main areas for consideration: defining terms; adopting a risk-based approach; clarifying the role of consent in relation to de-identification; and promoting codes of practice.

[1] For our introductory comments, we will use *de-identification* in its broadest sense – meaning any technology or process that reduces the identifiability of data – rather than to refer to the specific process of de-identification. We explore this further in the "Defining Terms" section of this submission.

***Defining Terms***

In the Discussion Paper, MGCS suggests that the Province of Ontario may "consider defining [de-identification] more clearly in law and setting clear guidelines for how privacy rules apply to [this type] of data."

CANON supports such an effort, and believes that defining terms will be critically important to clarify concepts and ensure common meanings that are well understood by all affected stakeholders. This includes organizations vis-à-vis their customers when describing their personal information management practices; controllers vis-à-vis processors when negotiating terms and conditions of a service agreement; and, organizations vis-à-vis regulators when demonstrating their accountability and data governance frameworks in the context of an investigation or audit. CANON has undertaken to develop a lexicon of key terms, having recently released the first component of this effort which speaks to the three states of data (identified; identifiable; and, non-identifiable). We would be happy to keep the MGCS apprised of this effort as it progresses.

With respect to defining terms, to the extent that they are helpful we offer the following suggestions:

- **Ensuring process neutrality**: The discussion paper makes specific reference to de-identification and its associated techniques, such as removing identifiers, obscuring information, or aggregating information. However, it is important to note that de-identification is one of multiple techniques which can be used to render information non-identifiable; others include data synthesis (in which a "synthetic" dataset is generated with the same statistical properties as the original) and homomorphic encryption (in which calculations are performed on encrypted data, without first having to decrypt it).

  We thus recommend that whatever definitions are adopted, they not set out highly specific criteria or prescriptive processes but rather focus on the end state that must be achieved, accommodating all potential means of rendering information non-identifiable. By doing so, organizations will be permitted to develop, and continually improve upon, robust and accountable processes using techniques which are rapidly evolving. This would avoid the risk of constraining innovation (or potentially lessening privacy protections) with definitions which in the near future may no longer be relevant or fit for purpose.

- **Consistency**: When defining terms, there is great benefit to looking broadly at other laws to ensure clarity and consistency of terms. Recognizing the reality that significant amounts of data flows (and will continue to flow) between sectors and across borders, it would be in Ontario's interest to develop definitions that are consistent, or at least interoperable, with those of other laws and other jurisdictions. This consistency would provide the level of predictability needed for organizations to design their current data practices, and to invest in new innovative ideas, with some level of assurance that they are grounded on well-established concepts.

- **Contextualization**: It has become clear, based on the collective experience of CANON members over many years, that the binary concept of personal information based solely on properties of the data itself is no longer fit for purpose. Identifiability is a relative concept that requires a contextual evaluation. It is therefore important that any definitions allow for consideration of

contextual factors such as: the nature of the data involved; the reasonable expectations of potentially affected individual(s); the intended purposes for its use; the release environment; the availability of other linkable data; the likely incentives to re-identify the data; the costs and level of expertise required to re-identify data; and, the potential harm to individuals should an individual be re-identified.

***Adopting a Risk-based Framework***

When considering the evaluation of non-identifiability, CANON members generally support the establishing of a risk-based framework.

We have seen such an approach successfully developed by many regulators, including the Ontario Information and Privacy Commissioner in its highly regarded guidance paper, De-identification Guidelines for Structured Data.[2] Risk-based frameworks have also been adopted as a globally accepted strategy within the statistical community.[3]

As part of this approach, CANON members also recommend that Ontario consider the adoption of a spectrum of identifiability rather than the existing black or white approach in which information is either identifiable or non-identifiable -- completely in or out of the ambit of privacy law -- respectively. For example, information that poses no serious risk of re-identification could remain outside of a private-sector privacy law, while information with a low risk of re-identification could be covered by the law, potentially exempted from consent (see below), but subject to other fair information principles as appropriate, including accountability, safeguarding and transparency. Such risk gradations would be determined in accordance with developed guidelines or standards and could be achieved using a variety of different technical methods and appropriate governance models.

Such a risk-based, spectrum approach to identifiability allows for a broad range of innovative uses of data, while accounting for, and mitigating, a reasonable amount of residual risk. It also allows for greater flexibility by allowing the use of other, innovative transformations of data, helping to encourage development of new privacy-enhancing technologies (a goal identified within the Discussion Paper) and future-proofing the legislation.

Ultimately, this would be more privacy-protective than a regime in which reasonable people could disagree on the bright line between identifiable and non-identifiable data, and which provides no protection for data which may fall in the grey zone in between. Moreover, such a risk-based approach is more conducive to innovative uses of data subject to appropriate protections. For instance:

- A highly detailed dataset, which would serve a significant public benefit but for which reidentification cannot be completely ruled out in the context of a public release, might instead undergo sufficient transformation to allow it to made available within a secure environment in which re-identification is meaningfully prohibited, while retaining its analytical value.

[2] https://www.ipc.on.ca/resource/de-identification-guidelines-for-structured-data/

[3] See, for example, Statistical Disclosure Control for Microdata: A Practice Guide, developed by a consortium led by the World Bank. https://sdcpractice.readthedocs.io

- Rather than performing analytics on raw data, an organization might create a de-identified dataset to provide to its internal analytics team and establish organizational and technical controls sufficient to mitigate against re-identification.

Such an approach can also serve as a vehicle towards increased accountability, as it would require organizations to create a robust governance framework for this "middle category" of data, which could include measures for identifying, documenting and mitigating risks. At the same time, a risk-based approach can allow organizations to focus on, and mitigate against, likely or significantly harmful risks, rather than spending limited resources attempting to exhaustively catalogue and mitigate against very remote and relatively harmless risks.

***Clarifying the Role of Consent in Relation to Data Rendered Non-Identifiable***

When considering the motivation to render data non-identifiable (including via de-identification), one of the most common challenges expressed by CANON members is a consistent understanding of the role of consent in relation to such data.

Ontario and the MGCS may thus wish to consider establishing a set of criteria by which it could be determined whether identifiable (i.e. data for which there is some reasonable chance of association with an identifiable individual) and/or non-identifiable data could be used or disclosed without consent. Such criteria may include consideration of socially- and economically-beneficial purposes that outweigh potential privacy harm resulting from possible re-identification.

We would further welcome clarification that the initial creation of de-identified or pseudonymized data which will be used or disclosed for those purposes does not require consent. We note on this point that, in many cases, an organization looking to create a de-identified dataset will not have a direct relationship with data subjects, nor have any means of contacting them to obtain consent to de-identify in the first place. Given the virtues of de-identification and other privacy-enhancing technologies as privacy protective measures which enables innovation, it would be a counter-productive result for organizations to have to collect and/or retain additional personal information for the purpose of protecting privacy in the future.

***Technical Standards and/or Codes of Practice***

Lastly, the Discussion Paper speaks to incentivizing the use of de-identification, potentially by supporting the creation of new technical standards. CANON agrees with this approach, but would recommend that it be expanded to include the development and recognition of Codes of Practice. We have seen the creation of Codes and frameworks in other jurisdictions, including the United Kingdom[4] and Australia[5] , and believe that a similar resource, developed through a stakeholder-initiated process and specific to the Canadian context, would be a highly effective way to promote the use of robust de-identification and other privacy-enhancing technologies by organizations in Ontario.

[4] United Kingdom Anonymisation Network (UKAN). "Anonymisation Decision-Making Framework" https://ukanon.net/ukan-resources/ukan-decision-making-framework/

[5] Commonwealth Scientific and Industrial Research Organisation (CSIRO) and the Office of Australian Information Commissioner (OAIC). "The De-Identification Decision-Making Framework." https://data61.csiro.au/en/OurWork/Safety-and-Security/Privacy-Preservation/De-identification-Decision-Making-Framework

CANON was created with just such an effort in mind. It is for all the reasons above that CANON has identified as one of its goals the development of a risk-based framework of principles to promote a more realistic approach to non-identifiability, subject to a robust governance model. We believe this proactive initiative will help support both private sector organizations seeking to innovate for economic and social prosperity, as well as public and health-sector institutions seeking to leverage their data for public good.

In the past year, we have also seen the creation of data protection certification mechanisms, seals and marks, largely encouraged by the incorporation of these concepts into the EU General Data Protection Regulation.[6] The introduction of a similar system of voluntary certifications in Canada may further incentivize privacy-protective operations within organizations. Such a system could, of course, be implemented in combination with a Code of Practice which would set out related technical and operations standards.

Our members look forward to the opportunity to engage with MGCS, technical experts, regulators, civil society, and/or groups such as the Standards Council of Canada. Collectively, we can work towards the creation of an accessible, operational resource that supports and enhances Ontario's position as a leader in both privacy protection and innovation.

***Conclusion***

Once again, the members of CANON appreciate this opportunity to provide feedback on MGCS's proposals for a private sector privacy law in Ontario, and look forward to have the opportunity to continue a discussion on the issues described above as well as the many benefits of the many techniques which can promote the privacy-protective use and sharing of non-identifiable data. Please feel free to contact Khaled El Emam or Vance Lockton, at info@deidentify.ca, any time.

Yours sincerely,

***The Canadian Anonymization Network***

[6] The General Data Protection Regulation (EU) 2016/679 (GDPR). Article 42.