CANON Canadian Anonymization Network

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

September 3, 2021

Dear Sir/Madam,

Thank you for the opportunity to provide comment on the Ministry of Government and Consumer Services (MGCS)'s *Modernizing Privacy in Ontario* white paper (the "White Paper").

We are writing on behalf of the Canadian Anonymization Network (CANON). 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. CANON'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 White Paper in its totality. Rather, the present submission is limited to comments about de-identification 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 White Paper and previous *Ontario Private Sector Privacy Reform 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 generation) as privacy-protective measures 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.

Our submission focuses mainly on the concepts of "anonymized" and "de-identified" information as those terms have been defined in the White Paper, as well as some broader comments regarding terminology more generally. These comments are in addition to the previous comments submitted in

October 2020 relating to the consultation *Reforming Privacy in Ontario's Private Sector* and associated *Privacy Reform Discussion Paper*.

Anonymized Information

The concept of "anonymized information", as defined in the White Paper, must be "altered irreversibly, according to generally accepted best practices, in such a way that no individual could be identified from the information, whether directly or indirectly by any means or by any person." We are concerned that this standard may be very difficult to achieve in practice.

In our internal discussions at CANON, our experts discussed one example that may potentially meet this standard: homomorphic encryption. Homomorphic encryption allows for analysis to be performed on fully encrypted data. In this case, the data could be considered "anonymized" under the White Paper's proposed definition because encrypted data is completely non-identifiable. Encryption seems at this time to be the only technology that could potentially meet this definition of "anonymized". However, encrypted data is always subject to possible decryption using a key that is designed for such purpose. Decryption must occur at some point in order to access the final results of any processing of the encrypted data. This raises the question of how can the key be sufficiently protected to meet the definition of anonymized information, since at some point the key must be accessed and used by man or machine to decrypt the information ?

The key may be split among multiple parties to ensure that no individual party has access to the full key. Even if the key is split among multiple parties, there is the possibility of collusion. The term "by any means" could be interpreted to mean that collusion among parties holding the key is possible and, under those circumstances, this approach would not meet the definition of anonymized information.

Furthermore, we note that encryption may not be the most practical means of safeguarding information as it introduces additional complexities to the handling of information.

If there is currently no practical way to meet this definition of "anonymization", we would encourage MGCS to consider whether it is useful to include such a definition in the legislation.

Further, and as discussed in more detail below, from an operational perspective, seeking to ensure consistency of terms across jurisdictions is important to provide the level of predictability needed for organizations to design their current data practices, and to invest in new innovative ideas. Insofar as the White Paper's proposed definition of "anonymization" conflicts with the definition of similar terms in other established regimes, this could create confusion and unnecessary complexity.

De-identified Information

Anonymized information is distinct from de-identified information in the White Paper. De-identified information is defined as "information about an individual that no longer allows the individual to be directly or indirectly identified without the use of additional information."

De-identified information here appears to be similar to the GDPR definition of pseudonymized data in that it is a reversible process and is therefore subject to certain privacy rights and requirements such as the implementation of a privacy management program and the need for security safeguards, as well as providing individuals with an opportunity to make a complaint or request information with respect to compliance. It is also exempt from certain obligations, which could be an incentive for organizations to apply these techniques.

We would like to acknowledge that "de-identified" as defined here could apply to many different levels of identifiability, from simple pseudonymization in which names and identification numbers are removed/replaced, to data that has had many transformations applied to it and, as a result, has a very low risk of re-identification. We want to propose that these variations ought to be taken into consideration in terms of how the requirements would be applied, and that the same level of protection would not be required for every data set regardless of the level of risk/identifiability.

As a result, we believe it would be clearer if the definition of "de-identified information" did not encompass pseudonymous information. Risk based de-identification requires the risk to be deemed acceptably low based on certain thresholds, for example, those included in the Information and Privacy Commissioner of Ontario's De-identification Guidelines for Structured Data. Pseudonymous data will not always meet those thresholds and, therefore, remains personal information. As such, more restrictions on the use and disclosure of pseudonymous data are required to both protect the data subjects as well as to provide an incentive for organizations to further de-identify their data. It would therefore be useful to draw a clear line between de-identified data and pseudonymous information, demonstrating that de-identification goes a step beyond pseudonymization, offering greater protection to data subjects. Consequently, we suggest that a definition of pseudonymous information, separate from de-identified information, be included and the obligations with respect to that form of information be clearly outlined. We are thus proposing that three levels of transformed data be defined: pseudonymous, de-identified, and anonymized (if that standard can be met).

We support the inclusion of a framework based on a risk-based approach to de-identification, and believe that these concepts will fit well within that framework. Also, we acknowledge the concept of proportionality, as included in the White Paper, which is defined in terms of the sensitivity of the data.

Terminology

Finally, we note that 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). Specifically, we believe the term "de-identification" is limiting in that it pertains to

only one method of rendering information non-identifiable. Other methods include data synthesis (in which a "synthetic" dataset is generated with the same statistical properties as the original data) and homomorphic encryption (in which calculations are performed on encrypted data, without first having to decrypt it), as well as future technologies that may be developed.

To revisit the comments we submitted in October 2021, we offer the following suggestions with respect to defining terms, to the extent that they are helpful:

- **Ensuring process neutrality**: The White Paper makes specific reference to de-identification and its associated techniques, such as removing identifiers, obscuring information, or aggregating information. As mentioned above, it is important to note that de-identification is one of multiple techniques that can be used to render information non-identifiable; others include data synthesis and homomorphic encryption.

  We recommend that the definitions that are ultimately adopted 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 that are rapidly evolving. This would avoid the risk of constraining innovation (or potentially lessening privacy protections) with definitions that may no longer be relevant or fit for purpose in the near future.

- **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 a significant volume 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 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.

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 we look forward to having the opportunity to continue a discussion on the issues described above, as well as the many benefits associated with techniques that promote the privacy-protective use and sharing of non-identifiable data. Please feel free to contact Khaled El Emam or Adam Kardash at info@deidentify.ca, should you wish to have a further discussion with CANON members on this important topic.

Yours sincerely,

The Canadian Anonymization Network
www.deidentify.ca