Treasury Board of Canada Secretariat
Secrétariat du Conseil du Trésor du Canada
Canada

Privacy Implementation Notice 2020-03: Protecting privacy when releasing information about a small number of individuals

Published: 2020-10-08

© Her Majesty the Queen in Right of Canada,
represented by the President of the Treasury Board 2020,

Published by Treasury Board of Canada, Secretariat
90 Elgin, Ottawa, Ontario, K1A 0R5, Canada

Catalogue Number: BT39-61/2020E-PDF
ISBN: 978-0-660-36870-2

This document is available on the Government of Canada website, Canada.ca

This document is available in alternative formats upon request.

Aussi offert en français sous le titre : Avis de mise en œuvre de la protection des renseignements personnels 2020-03 : protection des renseignements personnels lors de la diffusion de renseignements à propos d'un petit nombre de personnes

Privacy Implementation Notice 2020-03: Protecting privacy when releasing information about a small number of individuals

1. Effective date

This implementation notice takes effect on October 6, 2020.

2. Authorities

This implementation notice is issued pursuant to paragraph 71(1)(d) of the *Privacy Act*.

3. Purpose

This implementation notice provides guidance to institutions on measures to help protect privacy when releasing data about a small number of individuals, when the intention is **not** to release the identity of the individuals, in order to comply with the *Privacy Act*.

This notice sets out:

- suggested measures to protect against re-identification

- specific considerations when releasing information in response to access to information requests
- considerations when information about government employees or ministerial staff is proposed for release
- information sharing for administrative, research or statistical purposes

4. Context

Government institutions release data and information in various situations, such as:

- access to information requests
- audits
- evaluations
- statistical reports

When releasing data about a small number of individuals where the intention is **not** to release the identity of the individuals, there is a risk that one or more individuals could be identified in a dataset, even after names and other identifiers have been removed. This risk could occur through the information in the dataset alone or in combination with other sources of information. If there is a serious possibility that one or more individuals could be identified through the information, that information may constitute personal information as defined in section 3 of the *Privacy Act.* The disclosure of such personal information would constitute a breach of the privacy of those individuals (refer to Gordon v. Canada (Health), 2008 FC 258) unless the disclosure is otherwise authorized by the *Privacy Act* (refer to section 8).

To avoid privacy breaches through re-identification, government institutions need to reduce the risk of re-identifying an individual from the data, alone or in combination with other sources of information, to a level where there is no serious possibility of identification.

Once data has been successfully de-identified so that there is no serious possibility of re-identification, the data is no longer considered personal information and may be released, subject to any other restrictions, such as security considerations.

5. Guidance

The steps set out below can reduce the risk of re-identification. However, the risk of re-identification will not be entirely eliminated in all circumstances as it is not possible to know what other information is available to a data user. Note that the process for de-identifying data is more complex for datasets that have multiple variables. It is also important to be mindful of the heightened privacy risks associated with datasets that contain sensitive personal information, such as financial or health data. When considering the release of datasets that have multiple variables or that contain sensitive personal information, it is good practice to seek advice from statistical experts. It may be possible to consider approaches other than de-identification to mitigate privacy risks.

5.1. Measures to protect against re-identification

5.1.1. Suppress, mask or redact direct identifiers

Direct identifiers (also called directly identifying attributes or personal identifiers) are attributes that can be used alone to uniquely identify a data subject or point explicitly to an identifiable individual. Examples of direct identifiers are names, social insurance numbers and medical record numbers.

Direct identifiers generally add no analytical value and should be removed or replaced with a pseudonymous value.

5.1.2. Assess the risk of re-identification through indirect identifiers

Indirect identifiers (also called quasi-identifying attributes) are attributes that alone or in combination could be linked with other information, or used by someone who has background knowledge, to identify a specific individual. Examples of indirect identifiers are age, gender or province of residence.

If these attributes exist in the data to be released, institutions should assess the risk of re-identification by assessing:

- the fields within the dataset

- the information outside the dataset that could be linked to the indirect identifiers

The greater the number of indirect identifiers, the greater the likelihood (or possibility) of re-identification.

5.1.3. When releasing data in tables, determine the appropriate minimum cell size

A cell refers to data, usually presented in a summary table, about a group of individuals who share common attributes. A cell size (or count) is the number of individuals that share that unique attribute or combination of attributes. For example, if a dataset includes fields for sex and age, all records representing 40-year-old females constitute a cell. The cell size refers to the number of people in that cell.

By suppressing the information for cells that do not meet the minimum cell size, the risk of re-identification may not be eliminated, but it will be greatly reduced. Minimum cell size is closely related to the concept of a re-identification risk threshold. Therefore, totals for some columns or rows may need to be rounded, so that it is not possible to recover suppressed cells.

Re-identification risk can be expressed as the probability of correctly matching the identity of a data subject to their record. The probability of correctly identifying an individual is a factor of the chance of identifying the data subject from among the group. The risk of re-identification can therefore be calculated as 1 ÷ cell size. For example, a minimum cell size of 10 creates a re-identification risk of 0.1. As the minimum cell size increases, the risk of re-identification decreases.

Minimum cell size should be determined on a case-by-case basis, taking into account:

- the sensitivity of the data
- the potential harm that could result from re-identification
- the presence of vulnerable groups in the data
- the granularity of the data

- the age of the data
- the expectation of confidentiality or privacy of the data subjects when they provided the information
- whether the dataset contains all individuals in the real-world population of relevant data subjects (census), or whether the dataset is a sub-sample and how that sub-sample was selected
- whether the data collection was mandatory or voluntary
- whether the data will be released publicly, or whether access, use, and onward transfer of the data will be restricted through data-sharing agreements, governance policies or security controls
- the need to adopt a consistent approach if data releases are routine or regularly issued
- the existence of information already in the public domain that could be matched with the data and used to re-identify individuals

The information could be public as a result of previously released access to information requests, public reporting or public knowledge (such as when the information concerns a well-known figure).

There are also other approaches to setting an acceptable risk threshold or minimum cell size. These may be considered with appropriate expert guidance (refer to the Enquiries section of this notice).

Minimum cell size

As already noted, there is **no minimum cell size** that is appropriate for all data releases, and Treasury Board of Canada Secretariat policies do not specify a mandatory minimum cell size. However, the following best practices may serve as a starting point for a case-by-case analysis:

- A minimum cell size of 10 is often cited as a best practice for public data releases of data that is less sensitive, while a minimum cell size of 20 is cited for more sensitive data. This range is consistent with precedents identified in

other published guidance documents, such as the Information and Privacy Commissioner of Ontario's *De-Identification Guidelines for Structured Data*.

- When lawfully sharing data with trusted entities, such as other government departments, or through information-sharing agreements, lower minimum cell sizes may be acceptable. Alternative risk measures may be more appropriate in this context.
- From a practical perspective, the minimum cell size should never be less than three. This is because with a cell size of two, an individual who knows that they are a member of that cell could determine the identity of the other person in that category.
- Best practices can be found under "Other publications" in the References section of this notice.

5.1.4. Modify the data to mitigate the risk of re-identification

A number of techniques may be used to modify the data until there are no groups of individuals that have unique attributes or combinations of attributes below the minimum cell size. Techniques that can be applied include:

- suppressing (redacting) data for groups that do not meet the minimum cell size
- sampling data in order to share a representative dataset that is useful for analysis but that leaves uncertainty as to who is included from the broader population
- generalizing the values of the attributes in question in order to collapse small identifiable groups into larger ones (for example, age values could be generalized to age ranges, numbers could be rounded)
- using more sophisticated techniques to introduce unpredictability into the data (these techniques should be used only with expert advice)

After completing the measures to protect against re-identification, officials reviewing the dataset may determine that the information should not be released if the risk of re-identification is too great, or if data elements are suppressed to such an extent that the utility of the dataset is lost.

5.1.5. Document the process

Keep a record of the factors considered and techniques used to de-identify the data. It is important to be transparent with end users that the data has been de-identified to help them properly interpret the information. In addition, institutions should be prepared to demonstrate the broad steps that have been taken to protect privacy. However, the exact method of de-identification should not be released. Releasing the exact method may permit individuals to be re-identified by reversing the change or by combining the data with other information.

5.1.6. Plan for regular, ongoing and periodic assessment of re-identification risk

As the amount of information in the public domain increases and technological advancements continue, the factors used to assess and manage the risks associated with disclosure will need to evolve as well. It may be appropriate to revisit dissemination strategies or to amend the terms of information-sharing agreements as circumstances change.

5.2. Considerations for specific types of disclosures

Government institutions may release information through a number of different channels, some of which entail specific legal frameworks and considerations.

5.2.1. Responding to a request under the *Access to Information Act*

Section 19 of the *Access to Information Act* restricts the disclosure of any record that contains personal information, specifically records where there is a serious possibility that an individual could be re-identified. If information is suppressed (redacted) on this basis, the institution must:

- identify the exemption being applied
- notify the requester that they have a right to file a complaint with the Information Commissioner of Canada, consistent with sections 7 and 10 of the *Access to Information Act*

Exceptions to the prohibition on disclosure of personal information are set out in subsection 19(2) of the *Access to Information Act:*

- paragraph 19(2)(a) permits the release of personal information if the individual consents to the disclosure
- paragraph 19(2)(b) permits the release of personal information if it is publicly available
- paragraph 19(2)(c) authorizes disclosure of personal information if the disclosure is in accordance with section 8 of the *Privacy Act*

If the requested information falls under one of these exceptions, the personal information may be disclosed and there is no legal obligation to protect against re-identification. However, as the disclosure is discretionary, depending on the collection of the information or the context in which it was supplied, consideration should still be given to ensure that any privacy impacts are mitigated.

Depending on the circumstances, institutions may consider seeking consent from individuals in the dataset, especially in the case of small cell size.

If a record requested under the *Access to Information Act* contains personal information, altering the record through data de-identification techniques to protect privacy is not permitted. Instead, redaction of identifying aspects of the record is the appropriate approach to protect privacy. Care should be taken to mitigate the risk that the information that is released could be combined with other information released through access to information requests or with information from outside sources to enable re-identification.

5.2.2. Datasets containing information about employees of government institutions and ministerial staff

When releasing datasets that include information about government officers or employees, note that certain types of information are excluded from the definition of personal information (refer to paragraph 3(j) of the *Privacy Act*). Information that would normally be considered direct identifiers, such as name, title, or telephone number, may not need to be protected when related to the positions or functions of the individual, depending on the context of the request. However, if the excluded information appears in combination with indirect identifiers, the release of the information may enable protected personal information to be deduced.

For example, a dataset that contains only public servants' names and their salary ranges may be released as this information falls within the exception to the definition of personal information, provided that no additional indirect identifiers are found in or outside the dataset. However, if the dataset contains names and **exact** salaries, the information cannot be released as it is personal information. The application of subsection 19(2) may be considered.

Similarly, when releasing datasets that include information related to ministerial staff, certain types of information are excluded from the definition of personal information (refer to paragraph 3(j.1) of the *Privacy Act*). Information that would normally be considered direct identifiers, such as name and title, may not need to be protected when found in datasets. However, if the excluded information appears in combination with indirect identifiers, the release of the information may enable protected personal information to be deduced.

For example, a dataset that included only the names, titles and institution name of ministerial staff may be released, provided that no additional indirect identifiers exist. However, if the record or dataset includes information about their responsibilities or duties, that column could not be released as such information **is** personal information. The application of subsection 19(2) may be considered.

5.2.3. Information sharing for administrative, research or statistical purposes

Sharing personal information means that one or both parties may be disclosing to, or collecting personal information from, one another. Given the privacy implications, institutions should consider whether there are approaches to sharing personal information with other institutions that would improve privacy protection, even if both institutions are legally permitted to collect, use, retain or disclose the personal information. The following methods may be considered:

- sharing depersonalized information (removing all direct identifiers)
- sharing aggregated data (such as a range of ages instead of specific ages)

If the purposes of the information sharing cannot be achieved by using non-personal information, institutions should explore approaches to reduce or eliminate privacy

risks, such as:

- limiting use of the data to specific purposes
- limiting access to the data to certain individuals in secure workspaces
- limiting onward transfer of the data

When an institution receives personal information following an information-sharing agreement, it is responsible for ensuring security protection equal to what was required of the original data owner.

Information-sharing agreements for research and statistical purposes are provided for in paragraph 8(2)(j) of the *Privacy Act.*

Access to information and privacy coordinators and others should consult Treasury Board of Canada Secretariat's *Guidance on Preparing Information Sharing Agreements Involving Personal Information*.

6. Application

This implementation notice applies to the government institutions defined in section 3 of the *Privacy Act*, including parent Crown corporations and any wholly owned subsidiary of these corporations. However, this notice does not apply to the Bank of Canada or to information that is excluded under the *Privacy Act*.

7. References

Legislation

- *Access to Information Act*
- *Privacy Act*
- *Security of Information Act*

Related Treasury Board policy instruments

- *Directive on Open Government*
- *Policy on Access to Information*

- *Interim Policy on Privacy Protection*

Other publications

The following are comprehensive guidance documents on data de-identification:

- Office of the Privacy Commissioner of Canada: Draft Guidance on Data Anonymization and Disclosure Control (a GCaccount is required to view the draft guidance)
- Emam, Khaled El. *Guide to the De-Identification of Personal Health Information*. Boca Raton, Florida: CRC Press, 2013
- Information and Privacy Commissioner of Ontario: *De-Identification Guidelines for Structured Data*
- Information Commissioner's Office, United Kingdom: *Anonymisation: managing data protection risk code of practice*
- Australian National Data Service (ANDS): *ANDS Guide: De-Identification*
- Statistics Canada: Disclosure control (2014)
- Future of Privacy Forum: A Visual Guide to Practical Data De-Identification

8. Enquiries

Members of the public can email questions@tbs-sct.gc.ca for information about this implementation notice.

Employees of federal institutions may contact their access to information and privacy coordinator for information about this implementation notice.

Access to information and privacy coordinators may contact the Treasury Board of Canada Secretariat's Information and Privacy Policy Division at ippd-dpiprp@tbs-sct.gc.ca for information about this implementation notice.

For expert advice on how to assess the risk of re-identification, contact Statistics Canada at STATCAN.infostats-infostats.STATCAN@canada.ca.

▶ Report a problem or mistake on this page

Share this page

Date modified: 2020-10-08