De-Identification Guidance

Guidance Regarding Disclosure of Partially De-Identified and Anonymized Records

Corporate Information and Records Management Office

BRITISH COLUMBIA

October 2021 | Version 1

Table of Contents

INTRODUCTION

Purpose

The following guidance will support ministries' data and privacy professionals in determining when to de-identify personal information, to what extent to de-identify personal information, and what steps should be taken to ensure that de-identified records are disclosed lawfully and responsibly.

The objectives of this guidance are to:

- better enable meaningful information sharing and research, where there may be limited or no authority to share personal information in a given circumstance;
- encourage consistency in language associated with de-identification across government;
- reduce risks associated with the disclosure of inadequately de-identified records; and
- reduce the risk that a disclosure is not compliant under *Freedom of Information and Protection of Privacy Act* (FOIPPA).

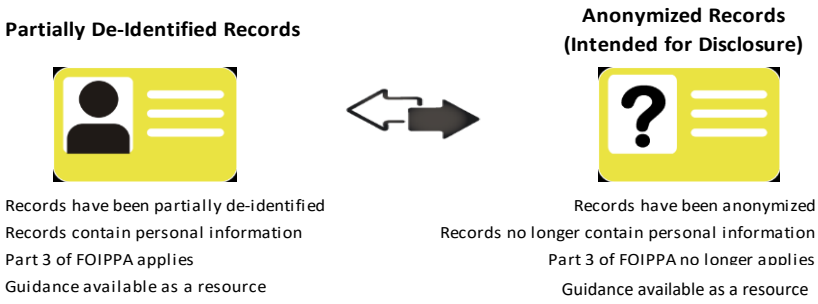
Background

De-identification is a process that includes the removal or transformation of personal information from a record to protect an individual's privacy. It is possible for de-identified records to be either (i) partially de-identified so that the individual may be re-identifiable; or (ii) anonymized so that the individual cannot be re-identified.

De-identification is one of many strategies a ministry may take when minimizing risk associated with disclosure of records. In order to determine which specific approach to de-identification may be best suited for a given record, it is helpful to start by considering whether the disclosure is authorized under Part 3 of FOIPPA.

- Disclosure Authority Available: When a disclosure of records that contains personal information is authorized by FOIPPA or other legislation, ministries may consider de-identifying (either partially de-identifying or anonymizing) personal information as a privacy protective strategy to secure information or mitigate risk.
- Disclosure Authority Otherwise Restricted: If ministries wish to disclose personal information where a compelling disclosure would otherwise be restricted under FOIPPA or other applicable legislation, ministries may choose to either anonymize or not disclose the personal information.

The below diagram distinguishes partially de-identified records from anonymized records and notes the respective FOIPPA Part 3 obligations:



This guidance does not instruct or dictate methodology on how to de-identify records; therefore, following this guidance will not result in confirmation that records are appropriately partially de-identified or fully anonymized. That decision requires subject matter expertise within the ministry.

In many situations, the processes involved with de-identification can be complex. With respect to anonymization, there will be at least some residual risk inherent to the disclosure of anonymized records. It is likely the case that ministry employees may need support from privacy and/or data professionals in their ministry or within the Office of the Chief Information Officer, particularly with more technical risk and/or de-identification analysis. This guidance supports ministries to make lawful and responsible decisions about partially de-identifying records and the *disclosure* of records that have been anonymized. Refer to Resources for additional support.

This guidance does not change or impact ministries' privacy impact assessment (PIA) or security threat and risk assessment (STRA) requirements. Ministries can contact their Ministry Privacy Officer or Ministry Information Security Officer for clarity on these requirements.

Application

This guidance was designed for use by ministries, but may be used by ministries, agencies, boards, and commissions subject to the Core Policy and Procedures Manual (referred to as ministries) when considering de-identification of data that is about an individual(s). The guidance is intended for use by data and privacy professionals.

Legal

These guidelines do not replace or limit a ministry's legal obligations under the *Information Management Act* or the *Freedom of Information and Protection of Privacy Act.*

Contact

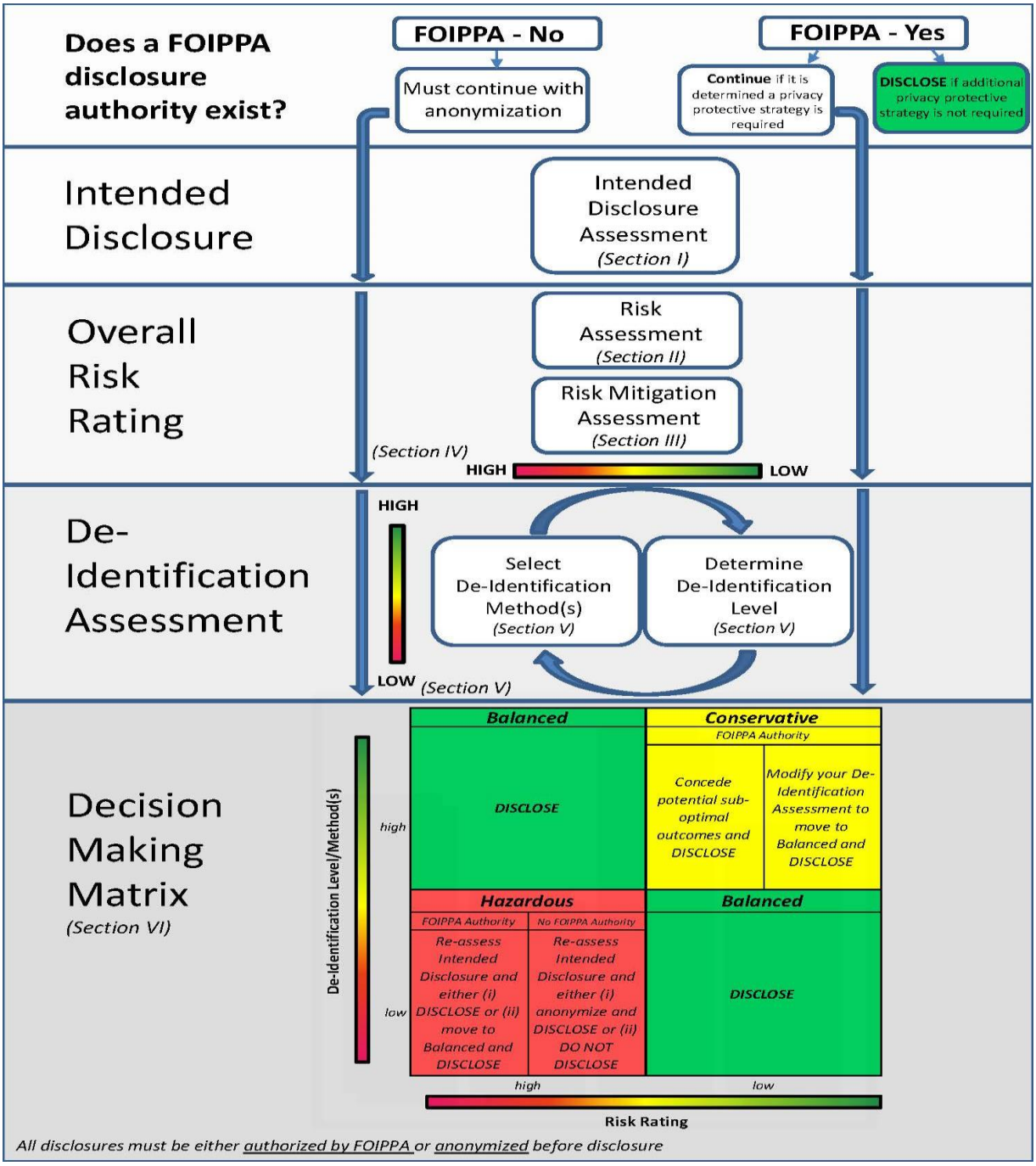
For questions or comments regarding these guidelines, please contact:

Privacy, Compliance and Training Branch
Corporate Information and Records Management Office
Ministry of Citizens' Services
Telephone: (250) 356-1851
Email: privacy.helpline@gov.bc.ca

GUIDELINES

De-Identification Decision Making Process

The following diagram is a guide for the de-identification decision-making process with comprehensive decriptions below. It is important to note that this process has been outlined sequentially; however, partially de-identified records or anonymized records are understood to be the product of an iterative assessment of the following steps:



1. Intended Disclosure Assessment

Firstly, ministries should determine if the intended disclosure is authorized by Part 3 of FOIPPA or other applicable legislation. If the disclosure is authorized, ministries should consider whether or not partial de-identification or anonymization is appropriate, e.g. would de-identification appropriately reduce risks associated with the disclosure. If the disclosure is not authorized, ministries would have to anonymize the personal information in order to disclose the record in accordance with FOIPPA.

Next, ministries should assess the business need or rationale for an intended disclosure of partially de-identified or anonymized records. The following factors should be assessed to support de-identification decision making.

1. Whether the intended disclosure is in support of data-sharing between ministries when authorities exist (e.g. there may be a strong business need from the ministries requesting the data and thus it may be prudent to take a corporate approach)?
2. What are the benefits associated with the intended disclosure (e.g. will public good result from the disclosure; will a disclosure provide benefit to private interest only)? [1]
3. What are the potential impacts of the disclosure (e.g. are there ethical implications that may cause undue harm to vulnerable or stigmatized populations)?
4. What are the risks associated with non-disclosure (e.g. is there a potential loss to the public good that may result from not disclosing the information)?

2. Risk Assessment

After an assessment of the intended disclosure, ministries should have an idea of the intended recipients, their intended use of the information, and other information that will assist in the next step – risk assessment. A risk assessment will assist with selecting the appropriate de-identification level and method(s) to support the intended disclosure. Risk level will depend on context (e.g. who is receiving the records and under what circumstances). Where the disclosure of de-identified information is ongoing, periodic re-assessment is advised.

Ministries should conduct the following risk assessments.

1. *Personal Information:* Ministries should assess factors such as the sensitivity, specificity, or uniqueness of the information; volume or complexity of the information; the mosaic effect; and risks associated with 'big data' (i.e. that re-identification may be easier because of the increasing abundance and availability of other information).
2. *Recipient, Environment, and Use:* Ministries should assess risks associated with the recipient, the disclosure environment *(e.g. online; secure data warehouse)*, and the subsequent use of the record after it is disclosed.
    a. Risks associated with the recipient: Ministries should assess the recipient.

[1] If there is a disclosure that is "(a) about a risk of significant harm to the environment or to the health or safety of the public or a group of people, or (b) for any other reason, clearly in the public interest", it may attract a requirement to disclose under FOIPPA. Please see guidance here for more information.

- For example, you may consider recipients that are subject to comparable privacy and access legislation and who have robust privacy or data governance as more "trusted" than others.

b. Risks associated with the disclosure environment: Ministries should assess factors related to the disclosure and the environment of disclosure, including whether or not the information will be published online and/or disclosed outside of Canada.
   - For example, there may be less risk associated with a secure data environment, given the built-in protections and controls on access to data, than with publication online.

c. Risks associated with the intended use: Ministries should assess what will happen to the record after it is disclosed, including potential risks associated with secondary use[2].
   - For example, there will be more risk associated with instances where the intended use occurs externally (where ministries can no longer control secondary usage) versus instances where the intended use occurs internally (where there are shared obligations and priorities).

3. *Re-Identification Risk:* Ministries should consider the possibility of, and potential motivations for re-identification (e.g. monetary gain, reputational harm, etc.). Ministries should also consider the resulting impacts including operational, financial, and legal, or personal harm to an individual or to a group (e.g. if it leaves an individual open to damage, distress, or financial loss, a more rigorous form of risk analysis and de-identification would be required).

Refer to Resources for additional information.

3. Risk Mitigation Assessment

De-identification is one disclosure control mechanism. The following tools may further reduce the risks identified above and support the legal and responsible disclosure of identifiable records, partially de-identified records, and anonymized records.

- Ministry-specific data governance, de-identification standards and/or documentation of de-identification processes.
- Understanding of roles and responsibilities for employees who i) conduct a de-identification risk assessment; ii) perform de-identification; iii) authorize the disclosure of anonymized records; and, iv) oversee the disclosure control process.
- Document de-identification process and results to demonstrate preventive efforts, risk-decision making and to support possible future audits or reported information incidents. A 'key' and/or 'crosswalk' resulting from the de-identification process that links back to the original data is an important step, and should be maintained exclusively in a secure environment.
- Use of audit tools to assess the recipients' compliance with any conditions associated with the de-identified information.
- Regular and ongoing re-identification risk assessments, including:
  - periodic review of de-identification methods;

[2] Secondary use of open data is encouraged, so any data that is made open should be appropriate for such use.

  - review that individuals performing de-identification have adequate and up-to-date training; and examination of the re-identification risk from the development of more sophisticated re-identification techniques, new information disclosures or potential collusion among data recipients.
- Further, where structured data is anonymized and is intended for disclosure to the internet; or to commercial organizations/entities, ministries should strongly consider:
  - implementing a process to periodically re-assess the anonymization of the data in consideration of increasingly sophisticated re-identification techniques (which may render anonymized data re-identifiable); and,
  - designating an employee responsible for overseeing the disclosure control process, including risk assessment, risk mitigation controls and re-assessment for re-identification risks.

Further to these, some organizations may choose to document the conditions and responsibilities surrounding the exchange of anonymized or partially de-identified records. These are not required by law or policy (except for Information Sharing Agreements, in some cases). Below are additional controls that may be appropriate when disclosing anonymized or partially de-identified records:

- A robust and secure disclosure environment (e.g. mature, secure public sector data warehouse).
- Entering into an agreement, as appropriate:
  a. Information Sharing Agreements (ISAs) or Research Agreements (RAs): If the exchange of records includes partially de-identified information, an ISA or RA may be required. See ISA guidance or RA resources for support in making this determination.
  b. Data Sharing Agreements: Data Sharing Agreements set conditions on disclosed data (e.g. restricting attempts at re-identification, intended use, intended disclosure medium, security controls, secondary uses or disclosures, etc.) where the information is considered anonymized. Ministries may use the ISA template and guidance as a model to support the development of a Data Sharing Agreement.
  c. User Access Agreements: If there is a desire to set conditions on the use of de-identified records at the user level, ministries may seek to do so with a one-to-one agreement.
  d. Implementing legal controls such as a Non-Disclosure Agreement or engaging in a General Services Agreement (GSA) to extend the obligations of FOIPPA to the service provider through the Privacy Protection Schedule (Schedule E).

4. Risk Rating

Overall Risk Rating: Ministries should evaluate the overall residual risk based on a balance of the above factors. A higher risk assessment in one of the above does not necessarily mean a higher overall risk assessment. For example, ministries who assess 'higher risk' on the spectrum based on the assessment of the personal information (e.g. sensitive health information) may be disclosing to a secure, trusted public sector environment and may further incorporate risk mitigation measures (e.g. information sharing agreements) in order to conclude that the overall risk rating is 'lower risk' on the spectrum.

Overall Risk Rating

Factors Contributing to Higher Risk	Factors Contributing to Lower Risk
more sensitive data	*less sensitive data*
publicly released	*limited access release*
recipient not covered under FOIPPA or comparable privacy and access legislation	*recipient covered under FOIPPA or comparable privacy and access legislation*
strong re-identification motivation could exist with high negative impact	*minimal re-identification motivation exists*
no agreements in place	*agreements in place*

Following the above assessments and overall risk rating, ministries may:

i. determine the outstanding risk is acceptable without further de-identification and disclose the records if authorized under Part 3 of FOIPPA (or other applicable legislation);
ii. determine that despite authorization under Part 3 of FOIPPA (or other applicable legislation), there remains outstanding risk to be addressed through de-identification and continue onto Section V: De-Identification Level/Method(s) Assessment; or,
iii. determine that a disclosure of information would not be authorized under Part 3 of FOIPPA (or other applicable legislation) and continue onto Section V: De-Identification Level/Method(s) Assessment to ensure information has been anonymized appropriately.

5. De-Identification Level/Method(s) Assessment

In order to proceed with the de-identification assessment, ministries should consider the appropriate (i) de-identification level; and (ii) de-identification method(s) through an iterative process. De-identification method refers to how you are making the information unidentifiable (as described below, e.g. suppression), and de-identification level refers to how robust you implement your de-identification method (e.g. suppressing fields with population under 5 subjects compared to suppressing fields with population under 50 subjects).

Ministries should select the de-identification level/method(s) that:

(i) either partially de-identifies or anonymizes the information (as appropriate);
(ii) considers the results of the Intended Disclosure Assessment above (e.g. increased de-identification level/method(s) may be applied in order to minimize the impact of a disclosure to a vulnerable or stigmatized population); and
(iii) are proportional to the overall risk rating, or higher (i.e. a higher risk rating indicates a need for higher de-identification level/method(s)).

The selected de-identification level/method will fall along the spectrum below (which includes some general characteristics that may assist with factoring in your decision).

De-Identification Level/Method(s)

Low De-Identification (partially de-identified records)	High De-Identification (anonymized records)
appropriate for limited access release	*appropriate for public data; DataBC*
increased data utility	*decreased data utility*
individual can be identified	*all direct and indirect identifiers have been de-identified (record can no longer be tied to an identifiable person)*

If a disclosure of records would not be authorized under FOIPPA, **the de-identification method(s) must be applied at a level that ensures the record is anonymized prior to disclosure.** In other words, proceeding with a disclosure of anonymized records where the records have not been anonymized with a de-identification level and method(s) that appropriately mitigates the risk of an individual being re-identified would not be authorized by FOIPPA.

Ministries may seek additional guidance from ministry-specific standards and processes or external subject matter experts on how to de-identify specific records.

The following list of de-identification methods is not exhaustive and more comprehensive guidance on de-identification methods can be found in Resources. Note that the risk of re-identification can be further reduced when methods are combined. Due to complexity, it may be necessary to seek assistance from experts in your ministry, or even to rely on software to de-identify personal information from a record.

Most commonly used methods of de-identification:

- Generalization – data manipulation in order to reduce the precision of the information (e.g. replacing date of birth with age groups, salaries with salary ranges, random digit rounding, etc.).
- Suppression – removing values or fields from the information (e.g. masking the personal information by removing identifiers such as name, address, date of birth, outliers, and unique demographics).

Other methods of de-identification that may be appropriate under specific circumstances:

- Aggregation – combining information such that the combined data reflects the attributes of a group rather than an individual (e.g. combining K-12 and post-secondary student information to reflect as a group for statistical analysis). It is possible that aggregate records may still contain indirect identifiers that may personally identify individuals (e.g. small populations).
- Disaggregated Data – refers to data broken down by age, gender, sex, race, ethnicity, income, education, etc. This is sometimes referred to as a sex- or gender-disaggregated data. Similar to aggregation, disaggregated data is de-identified data that may be partially de-identified or anonymized.

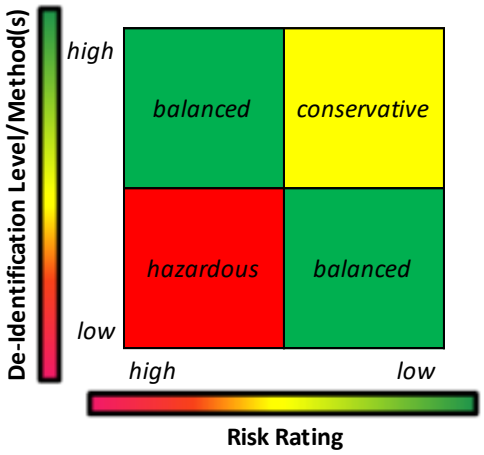
- Subsampling – disclosing a random subset of the record rather than the full record containing all of the personal information (e.g. national statistics agencies releasing only a small subsample of census data).
- Randomization – replacing actual values with random values that look real (e.g. replacing real names and addresses with simulated names and addresses).

6. De-Identification Decision Making

To assist ministries in determining the appropriate de-identification level/method(s), the table below illustrates the trade-off between risk and approach. Based on the assessments completed above, ministries may plot the overall risk rating determined (low/high on the x axis) and the de-identification level/method(s) selected (low/high on the y axis). Ministries will then find themselves in one of four possible quadrants assessed as either: Conservative, Hazardous, or one of two Balanced quadrants.

The level of de-identification is not fixed for all intended disclosures. In order to be "balanced", the same information may involve different levels of de-identification depending on the intended disclosure and the overall risk. For example, ministries may provide partially de-identified information to smaller regional districts but may require higher de-identification level/method(s) for that same information to be distributed online.

Decision Making Matrix



Balanced: Ideal State. Note that a disclosure of partially de-identified records in a balanced quadrant must always be authorized under Part 3 of FOIPPA.

Conservative: Instances where risk rating is low and de-identification level/method(s) is high. If ministries are concerned about moving from a conservative to a balanced approach, they may wish to return to the risk assessment to ensure that the risk rating appropriately reflects concerns. If ministries find themselves in the conservative quadrant, they should:

- review the intended disclosure;

- incorporate risk mitigation measures to move into a balanced quadrant; or
- concede potential sub-optimal outcomes, such as loss of data utility, increased bureaucracy, and/or delays to valuable insight of the intended disclosure and maintain a conservative approach. For example, without geographical information (region, city, etc.), information cannot be segmented to find regional patterns or differences.

Hazardous: Instances where risk rating is high and de-identification level/method(s) is low indicate that the assessed risk is not being effectively managed. Operating in this quadrant can have significant negative legal, financial, and regulatory consequences for ministries.

The disclosure of anonymized information must never be in the hazardous quadrant because high de-identification level/method(s) is required in order to anonymize the information. Additional disclosure controls must be implemented where outstanding risk mitigation is deemed necessary (refer Risk Mitigation Assessment for further detail).

The disclosure of partially de-identified information may persist when categorized as hazardous where the disclosure is authorized under FOIPPA or another enactment (i.e. where the disclosure of personal information would be authorized under FOIPPA). However, additional disclosure controls should be considered (refer Risk Mitigation Assessment for further detail).

Ministries whose assessments result in a possibility of operating in the hazardous quadrant should complete the following:

- Re-evaluate the intended disclosure to determine if there is a positive business need or interest served.
- Ministries should not proceed if the disclosure would not be authorized under FOIPPA (or other applicable legislation). If the disclosure of records is authorized under FOIPPA (or other applicable legislation), the information may be disclosed. Remember that the authority to disclose is generally discretionary (excluding where information sharing is obligated under legislation).
- Incorporate measures to move into a balanced quadrant by lowering the overall risk rating (e.g. adding risk mitigation measures) and/or by increasing de-identification level/method(s). If a disclosure of records would not be authorized under FOIPPA, the resulting de-identification level/method(s) must result in anonymized records.

7. Re-Identification Response Process

The privacy protections in FOIPPA do not apply to information that cannot personally identify an individual. Ministries should take an active role in securing de-identified information from either becoming personal information (e.g. combining multiple de-identified datasets that inadvertently now identify an individual) or from a re-identification attack.

Should any government employee discover an actual or suspected incident where de-identified information may have been re-identified, and is not authorized by FOIPPA, the employee must comply with the Information Incident Management Policy and immediately notify their supervisor and report

centrally to CIRMO and the Office of the Chief Information Officer (OCIO) via a (toll-free) dedicated phone line: 250-387-7000 (toll-free: 1-866-660-0811) and select option 3.

DEFINITIONS

Aggregation is one method of de-identification that combines information so that it reflects the attributes of a group rather than an individual. It is possible that aggregate records may still contain indirect identifiers that may personally identify individuals (e.g. small populations).

Anonymization is a de-identification process that removes or transforms all direct and indirect identifiers in a record for which there is a reasonable expectation that the identifiers could be used, either alone or with other information, to identify an individual. An anonymized record no longer contains personal information; therefore, the privacy protection provisions contained in Part 3 of FOIPPA or other applicable legislation no longer apply.

Big Data refers to the amount of data available and the use of advanced data analytics to extract valuable information from the data.

De-Identification is a process that removes or transforms direct and indirect identifiers in a record. De-identification methods can include generalization, suppression, aggregation, randomization, etc. De-identification methods for unstructured data can include redacting or severing, etc. De-identification processes may result in (i) partial de-identification; or (ii) anonymization.

Direct Identifiers identify an individual without additional information (e.g. name, address, telephone number). Direct identifiers are personal information.

Disaggregated Data refers to data broken down by age, gender, sex, race, ethnicity, income, education, etc. This is sometimes referred to as sex- or gender-disaggregated data. Disaggregated data is de-identified data that may be partially de-identified or anonymized.

Indirect Identifiers may identify an individual when they are connected with other pieces of information to single out an individual (e.g. age, gender, date of visit). While indirect identifiers on their own may not be personal information, they are considered personal information if they can be combined together to identify an individual. This is commonly referred to as the mosaic effect.

Partial De-Identification is a de-identification process that removes direct identifiers and may manage the indirect identifiers that could potentially be combined to identify an individual. Partially de-identified records contain personal information; therefore, the disclosure of partially de-identified records would require appropriate authorization under Part 3 of FOIPPA. Some partially de-identified information may also be referred to as pseudonymized data.

Personal Information is, according to FOIPPA, recorded information about an identifiable individual other than contact information.[3]

Re-Identification is the reversal of de-identification. Re-identification encompasses any process of re-establishing the link between the de-identified data and the individual the information is about.

Structured data is data that is machine-readable in a fixed field within a record or file (e.g. a database or spreadsheet), as opposed to unstructured data (e.g. a report in pdf format).

[3] Contact Information, according to FOIPPA, means information to enable an individual at a place of business to be contacted and includes the name, position name or title, business telephone number, business address, business email or business fax number of the individual.

RESOURCES

Government Resources

1. See your ministry intranet site for contacts related to data quality, data custodians, data stewards, and/or business intelligence
2. Ministry Privacy Officer
3. Ministry Information Security Officer
4. Privacy and Access Helpline