



Health and Social Services De-Identification Policy

Policy Statement

A consistent approach must be used across the health and social services system when applying de-identification methods to comply with safeguard requirements and ensure personal health information is protected.

Scope

This Policy applies to all employees of the Department of Health and Social Services (Department) and the Health and Social Services Authorities (HSSAs) who are tasked with de-identification.

Definitions

The following terms apply to this Policy:

"De-identification" refers to the process of manipulating personal health information so that the identity of the individual(s) the information is about cannot reasonably be determined.

"Employees" for the purpose of this Policy means all individuals employed by the Department or HSSA including information managers.

"Health and Social Services Authorities" refers to the Northwest Territories Health and Social Services Authority and Boards of Management under the *Hospital Insurance and the Health and Social Services Administration Act*.

"Information manager" means a person or organization that provides one or more of the following services for the Department or HSSA:

- (a) the processing, storage, retrieval or disposal of personal health information,
- (b) the transforming of personal health information, including the transforming of personal health information to create or produce non-identifying information,
- (c) information management services, information system services or technology services.

"Manipulating" refers to, stripping, encoding or otherwise transforming personal health information in the context of personal health information.

"Non-identifying information" refers to personal health information that cannot be reasonably used to determine an individual's identity or lead to re-identification.

"Personal health information" means information in any form that identifies an individual, or in respect of which it is reasonably foreseeable in the circumstances that the information could be used, either alone or with other information, to identify an individual:

- (a) information about the health and health care history of an individual,
- (b) information respecting health services provided to an individual,
- (c) information about eligibility or registration of an individual for a health service or related product or benefit,
- (d) information about the payment for a health service for an individual,
- (e) information collected in the course of providing a health service to an individual or information that is collected incidentally to the provision of a health service to an individual, including the individual's name and contact information,
- (f) a personal health number, other identifying number, symbol, or other particular assigned to an individual in respect of health services or health information,
- (g) prescribed information about a health service provider that provides a health service to an individual,
- (h) information respecting the donation by an individual of a body part or bodily substance,
- (i) information prescribed as personal health information in the Health Information Regulations.

"Re-identification" means the process of manipulating previously de-identified personal health information so that the information can reasonably identify an individual or individuals the information is about.

Provisions

- 1) The Department/HSSAs must
 - a. take measures to protect personal health information when applying de-identification methods; and,
 - b. ensure de-identification methods performed by employees meet the standards and best practices as set out in the Personal Health Information De-Identification Guidelines (Appendix 1), as amended from time to time.
- 2) De-identification methods may only be applied by authorized employees, identified by the Department or HSSAs.
- 3) The DHSS Process for Personal Health Information De-Identification (Appendix 1) must be followed when applying de-identification methods.

- 4) Authorized employees must consider the following when determining the appropriate methods of de-identification:
 - a. who may access the de-identified information;
 - b. why the de-identified information may be accessed, and,
 - c. the risk of re-identification, including consideration of whether there is other information available that could lead to re-identification.

Authority and References

Access to Information and Protection of Privacy Act

Health Information Act

Ministerial Directive 2016-02: Privacy Standards, Policies and Standards

<original signed by>

Deputy Minister

May 3, 2017

Date

Personal Health Information De-Identification Guidelines

As employees of the Health and Social Services system, it is an obligation to ensure we protect client information when using health data and in particular when applying de-identification methods.

The Department has developed these guidelines¹, to assist Department and HSSAs employees in:

- applying appropriate de-identification methods
- assessing the risk of re-identification.

Definitions

"Aggregate-level data" means data collected at the record-level then have been averaged or grouped into ranges and reported as a sum to ensure there are no directly identifying variables.

"De-identification" means the process of manipulating personal health information so that the identity of the individual or individuals the information is about cannot reasonably be determined, making the information de-identified health information, the use of which either alone or with other available information cannot reasonably lead to re-identification.

"Record-level data" means data at the level of an individual person.

"Re-identification" means the process of manipulating previously de-identified health information so that the information can reasonably identify an individual or individuals the information is about.

Applying de-identification methods

Before you apply any de-identification method:

- Be aware of de-identification principles which include:
 - Disclosures are made with the minimum amount of data.
 - A number of de-identification methods are commonly used and available for application (see *Recommended de-identification methods*).
 - No single method can independently meet the diverse needs of de-identifying personal health information.
 - Each de-identification project or file requires a unique process in terms of type of de-identification method to be applied and combination of methods.
 - Be capable of applying current de-identification tools and understand the statutory requirements related to de-identification.

To choose an appropriate de-identification method:

¹ These guidelines have been developed with permission to use and adopt information from *Pan-Canadian Deidentification Guidelines*, produced for the Office of the Privacy Commissioner of Canada, April 2007, the Tools for De-identification of Personal Health Information prepared for the Pan Canadian Health Information Privacy Group, September 2009, and the *Best Practice Guidelines for Managing the Disclosure of De-identified Health Information*, produced for the Canadian Institute for Health Information, October 2010.

- Review the de-identification request and decide which level of data to be released - record-level vs. aggregate data. Record-level data are more prone to re-identification than are aggregate data.
- Decide which method(s) is most appropriate. Data reduction methods are more commonly used than data modification. Data modification methods involve more radical approaches to the data and have greater potential to reduce the informativity of the data.
- Begin with Reduction in Detail method followed by Suppression. These are:
 - the most accepted methods in practice,
 - the least expensive to apply,
 - the easiest to understand, and
 - the easiest to predict re-identification risk.
- When you apply the pseudonymization method, ensure the generated pseudonyms are specific to a given data set. It protects control of the data and future capability of data linkage, if needed.

Assessing re-identification risk levels

The purpose of risk assessment is to determine how much de-identification to perform in order to reduce the risk of re-identification to an acceptable level. (See *Dependency Table*)

The assessment of re-identification risk levels may include:

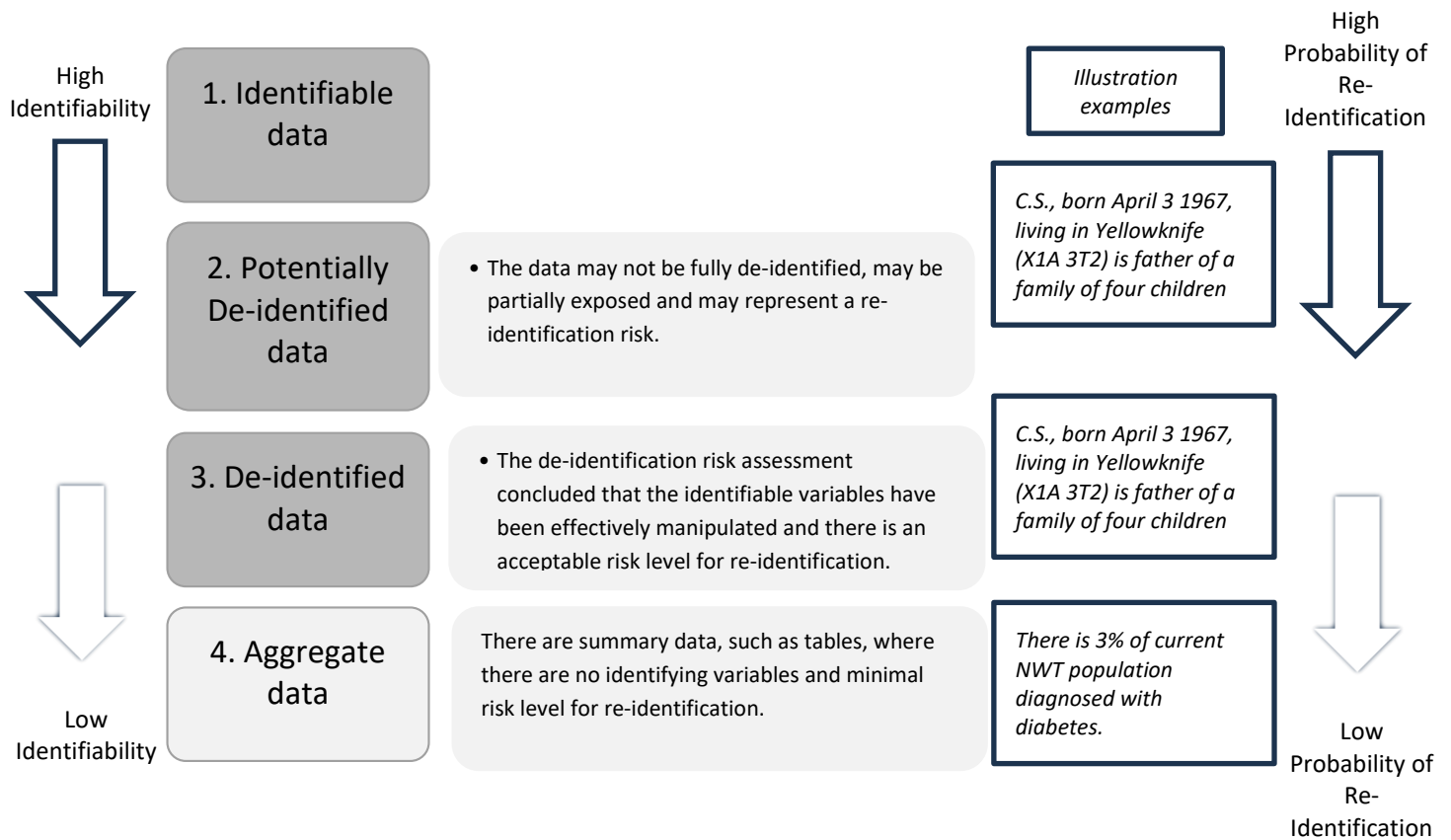
- Map the risk level of re-identification in order to ensure appropriate de-identification methods application. Refer to *2011 CiHi Best Practice Guidelines for Managing the Disclosure of the De-identified Health Information* and *2007 Pan-Canadian De-Identification Guidelines* documents for further assistance.
- Assess intention and capacity to re-identify; mitigating controls, estimate probability of a reidentification attempt, estimate how much de-identification is required.
- Consider establishing flexible guidelines for acceptable levels of re-identification risk that can address a range of re-identification risk scenarios.

The risk of re-identification may shift as technologies develop and a larger amount of information is available. Re-assessment of the re-identification risk levels on a regular basis is advised, and may include:

- Assess if higher level of de-identification is needed;
- Assess if a release of additional de-identified information could potentially assist in the re-identification of already available information.
- Assess the features of the analytics reporting in e-systems when querying an e-system.

Dependency Table

Identifiability and Probability of re-identification based on different types data



Recommended de-identification methods

De-identification methods		Application		
		Geo-Code	Numeric	Alphanumeric
Data Reduction	Reduction in Detail Involves reduction in the detail of the data through rounding or collapsing the data values into larger categories.	Reduce postal codes to first 3 characters, i.e. Forward Sortation Area.	Round birth dates to year. Express dates relative to a milestone date.	
	Suppression Involves the removal or withholding of the data records.	As a rule of thumb, suppress geo-codes when they contain six observations or less.	As a rule of thumb, suppress numbers when they contain six observations or less.	As a rule of thumb, suppress alpha variables when they contain six observations or less.
	Sampling Involves random samples of the records from a database.	E.g. Database contains too many records for analysis. If analysis was conducted using a random sample or one-fifth of the records from a database and additional technique was applied (e.g. reduction in detail). Sampling is often used in combination with other techniques, rather than on its own to protect against risk of re-identification.		
Data Modification	Random addition of 'noise' Involves adding random 'noise' to the values of data in order to disguise its true value.	Randomly change the actual postal codes within certain predetermined acceptable range (e.g. first or last two)	Randomly change the actual date of birth within certain pre-determined acceptable range (e.g. 6 months)	Randomly change the name within certain pre-determined acceptable range (e.g. first or last)
	Data swapping Involves identification of records of pairs of individuals with roughly the same characteristics and the data values are then swapped between the two records.	De-identified personal health information is created with records that are no longer the original records, but the analysis produces same results as achieved using original personal health information.		
	Substitution Removes the association between the individual and the identifying data by replacing original data with random or exchanged values.	If postal code is manipulated then ensure telephone area code is consistent.	If health card number is manipulated then ensure the new number can pass a checksum validation check.	Select new names in same proportion as in general public If surname is manipulated then ensure the new name has the same number of characters and ethnicity.
	Pseudonymization Removes the association between the individual and the identifying data by replacing the data with one or more pseudonyms (aka codes).	Can be applied to most geo- data.	Can be applied to most numeric data. Date of birth (and/or name) can be replaced with a unique pseudonym. Pseudonyms should be independently generated.	Can be applied to most alpha data. To facilitate linkages across databases, the pseudonym generated for the same individual must be consistent.