**Types of Data**

Researchers may seek to collect, use, share and access different types of information about participants to meet a research objective with Research Ethics Board approval. Such information may include personal characteristics or other information about which an individual has a reasonable expectation of privacy (e.g., age, ethnicity, educational background, employment history, health history, life experience, religion, social status).

This document describes identifiable and non-identifiable information as defined in [TCPS2](https://ethics.gc.ca/eng/tcps2-eptc2_2022_chapter5-chapitre5.html):

***Identifiable Information:***

*Directly identifying information*

The information identifies a specific individual through direct identifiers (e.g., name, social insurance number, personal health number).

If datasets or other data sources accessed for this study will contain identifiable information and for the purpose of data analysis the following should be included in the data management plan.

* Data captured on data collection tools will include direct and indirect identifiers. These include e.g., name, email, address, phone number, date of birth, partial date of birth, medical record number, pathology number, full postal code, social insurance number.

*Indirectly identifying information*

The information can reasonably be expected to identify an individual through a combination of indirect identifiers (e.g., date of birth, place of residence or unique personal characteristic).

If datasets or other data sources accessed for this study will contain identifiable information and will be de-identified for the purpose of data analysis the following should be included in the data management plan.

* Data captured on data collection tools will include indirect identifiers. These include e.g., date of birth, partial date of birth, full postal code.

*De-identified/Coded information*

Direct identifiers are removed from the information and replaced with a code. Depending on access to the code, it may be possible to re-identify specific participants (e.g., the principal investigator retains a list that links the participants’ code names with their actual names so data can be re-linked if necessary).

If datasets accessed will include direct identifiers but the information will be de-identified, include the following in the data management plan.

* Datasets accessed for this study will contain indirect identifiers. These include e.g., date of birth, partial date of birth, medical record number, pathology number, full postal code. No direct identifiers will be included in datasets.

***Non-identifiable information:***

*Anonymized information*

The information is irrevocably stripped of direct identifiers, a code is not kept to allow future re-linkage, and risk of re-identification of individuals from remaining indirect identifiers is low or very low. If datasets accessed will include anonymized data, include the following in the data management plan.

* Datasets accessed for this study will contain indirect identifiers. These include [e.g., date of birth, partial date of birth, sex, gender, partial postal code]. No direct identifiers will be accessed or collected for this study.

*Anonymous information*

The information never had identifiers associated with it (e.g., anonymous surveys) and risk of identification of individuals is low or very low. If datasets accessed will include anonymous information, include the following in the data management plan.

* All data collected will be anonymous. No direct or indirect identifiers will be accessed or collected for this study.