DALHOUSIE UNIVERSITY Health Data Nova Scotia **De-identification and Linkage Policy**	**Author:** S.Kennedy	**Review Date:** 01.01.2018
	Approved by and date: S.Carrigan / 05.04.2017	**Effective Date**: 05.04.2017
	Version Number: v1.0	***Page 1 of 5***

1. BACKGROUND & PURPOSE

1.1 The purpose of this policy is to set out the protocol used for:

- Reviewing data when received from the Nova Scotia Department of Health and Wellness (DHW) to ensure that intended direct identifiers are removed and that Unique Identifiers such as Health Card Numbers, Provider Numbers, and Social Insurance Numbers are encrypted.
- Performing linkages between datasets held at Health Data Nova Scotia (HDNS); and
- Performing linkages between data provided by a researcher and de-identified HDNS data.

2. APPLICATION

2.1 This policy applies to researchers and health service assessment analysts who request access to data through HDNS - referred to as requestor(s) - and HDNS staff and contractors.

3. DEFINITIONS

3.1 *Administrative Data:* Information collected primarily for administrative (not research) purposes. These data are collected by government departments and other organizations for the purposes of registration, transaction and record keeping, usually during the delivery of a service.

3.2 *Aggregate Level Data*: The result of applying statistical procedures (e.g., weighting, imputation) and analyses (e.g., means, regression) to individual-level data sets.

3.3 *Data Access Committee (DAC)*: The Committee tasked with reviewing requests to conduct secondary data analysis for research or health service assessment purposes using the administrative databases held by HDNS for privacy, security, and confidentiality concerns.

3.4 *Data Linkage*: The bringing together of two or more records of personal health information to form a composite record.

3.5 *De-identified Information*: Information that has had all identifiers removed that

(i) identify the individual, or

(ii) could be reasonably foreseen to be utilized, either alone or with other information, to identify the individual.

3.6 *Encryption*: The process of obscuring information, often through the use of a cryptographic scheme or algorithm in order to make the information unreadable without special knowledge; i.e. the use of code keys.

3.7 *External Analyst*: any statistical analyst operating outside of HDNS.

3.8 *Personal Health Information:* Identifying information about an individual, whether living or deceased, and in both recorded and unrecorded forms, if the information:

(i) relates to the physical or mental health of the individual, including information that consists of the health history of the individual's family,

(ii) relates to the application, assessment, eligibility and provision of health care to the individual, including the identification of a person as a provider of health care to the individual,

(iii) relates to payments or eligibility for health care in respect of the individual,

(iv) relates to the donation by the individual of any body part or bodily substance of the individual or is derived from the testing or examination of any such body part or bodily substance,

(v) is the individual's registration information, including the individual's health-card number, or

(vi) identifies an individual's substitute decision-maker.

3.9 *Unique Identifiers*: Information that can directly identify an individual and includes: name, street address and identifying numbers (e.g., health card number, physician identification number, employee ID, Social Insurance Number).

4. POLICY STATEMENT

4.1 HDNS only stores personal health information in the form of administrative data that has had Unique Identifiers removed prior to receipt by HDNS.

4. 2 Unique Identifiers are encrypted by data providers prior to data being provided to HDNS, using an algorithm unknown to HDNS. HDNS does not accept transfer of or allow linkage to external datasets containing unencrypted Unique Identifiers.

4.3 When providing linkages between data sets, HDNS minimizes the risk of identification of individuals by only allowing access to a minimal dataset and by requiring the suppression of small cells (<5) in the reporting of results.

5. PROCEDURES

5.1 Data received from DHW and Medavie Blue Cross

5.1.1 Medavie Blue Cross and DHW encrypt the Unique Identifiers attached to any data to be provided to HDNS, using their encryption algorithm, and send this data with the encrypted numbers to HDNS.

5.1.2 Upon receipt, the HDNS Senior Data Analyst or the Data Documentation Specialist review the data to ensure that all the Unique Identifiers are encrypted and that the variable names and attributes are consistent. If there is concern about the encryption, HDNS will contact the data provider to notify them of the issue and will delay using the data until the issue is addressed.

5.2 Internal Data

5.2.1 To prepare datasets for access, HDNS extracts only the Data Access Committee (DAC)-approved variables from the database, performing any required individual-level data linkage using the encrypted Unique Identifier(s). The encrypted Unique Identifier(s) is then removed and replaced with a study ID. Wherever possible, flags or derived variables (e.g., a yes/no for a group of conditions rather than providing a specific diagnostic code) are used.

5.2.2 The dataset is project specific and access is permitted for only the analyst for that project.

5.2.3 External Analysts have two options to access the data: using remote secure access to the HDNS Secure Data Platform, or accessing a secure HDNS workstation.

5.2.4 When using the HDNS Secure Data Platform, the External Analyst logs on to password-protected secure account where only data for which access by that analyst are located.

5.2.5 If an HDNS Analyst is performing the analysis, only aggregate level data (with cells less than 5 suppressed) is released to the project team in the form of statistical analytic output, tables, graphs or charts.

5.2.6 Linked data sets are only maintained on the HDNS Secure Data Platform as required by the HDNS guidelines or REB requirements. The data is then destroyed in accordance with the **Data Retention, Destruction and Restoration Policy.**

5.3 External Data

5.3.1 Requestor(s) who wish to link external datasets with HDNS datasets must ensure that any data sent to HDNS uses encrypted Unique Identifiers and is the minimal amount of data required for the project.

5.3.2 If the other datasets contain unencrypted Unique Identifiers, the requestor(s) must send the unencrypted Unique Identifiers and the HDNS study ID to Medavie Blue Cross for encryption of the Unique Identifiers.

5.3.3 Medavie Blue Cross sends the encrypted Unique Identifiers and study ID to HDNS.

5.3.4 The requestor(s) send the data file with study ID and approved variables only to HDNS.

5.3.5 HDNS creates the project dataset (consisting only of the approved variables/flags) for those encrypted Unique Identifiers per data access and REB approvals, removes the encrypted Unique Identifiers from the study dataset and attaches the study IDs.

5.3.6 HDNS links the datasets by study ID and stores them in a project specific, password-protected, monitored account on the HDNS Secure Data Platform.

5.3.7 The requestor(s) are then able to access the linked project dataset on the HDNS Secure Data Platform or HDNS workstation.

6.0 ADMINISTRATION

6.1 Accountability

6.1.1 The Senior Data Analyst and Data Documentation Specialist are responsible to ensure that all data received from data providers do not contain direct or indirect identifiers.

6.1.2 HDNS Analysts are responsible, when analyzing data, to ensure that cell sizes less than 5 are not released and that data are only released in aggregate form.

6.1.3 Requestor(s) are responsible for ensuring external datasets sent to HDNS do not contain unencrypted Unique Identifiers other than those approved by the DAC.

6.2 Monitoring, Auditing and Reporting

6.2.1 The study datasets and/or analysis results are vetted by the HDNS Senior Data Analyst or System Administrator from a privacy perspective before being released to the requestor(s).

7. RELATED POLICIES AND OTHER DOCUMENTS

7.1 HDNS Policies and Procedures

- Research Plan Policy
- Data Access and Confidentiality Policy
- Data Retention, Destruction and Retention Policy
- Security Policy

7.2 HDNS Forms

- **n/a**

7.3 Other Documents

- **n/a**