DALHOUSIE UNIVERSITY
Health Data Nova Scotia

Data Access Request Instructions

OVERVIEW

To obtain data access approval for research or health service assessment projects, requestors must complete our Data Access Request Form and submit supporting documentation (i.e., project proposal, research ethics board (REB) submission (if applicable), CV, etc.). This instruction document contains important information you will need to ensure a successful data request, including an overview of how the Data Access Process works, ethics and privacy requirements and definitions and detailed steps for completing and submitting your request.

ACRONYMS

DAC Data Access Committee

DAR Data Access Request

DHW Nova Scotia Department of Health & Wellness

HDNS Health Data Nova Scotia

PHIA Personal Health Information Act

REB Research Ethics Board

HDNS DATA ACCESS COMMITTEE

Your request for data from Health Data Nova Scotia (HDNS) will be reviewed by the HDNS Data Access Committee (DAC).

The responsibilities of the DAC are to:

- Ensure the data requested shall be appropriate and specific to the research objectives and methodology.
- Protect the confidentiality of personal health information in the custody of HDNS and the privacy of the individual who is the subject of that information.
- Ensure that conditions pursuant to the HDNS agreement with the Department of Health and Wellness (DHW) for access to DHW data have been met.
- Uphold standards of data access consistent with the highest levels of security, confidentiality and privacy in Canadian legislation, namely:
    - To maximize the protection of individual privacy
    - To approve access to linked data files only to nominated researchers involved in specific, approved research projects
    - To approve access by researchers to minimum datasets required for their specific project
    - To provide data to support approved quality review initiatives, and
    - To assure data custodians that those data which are their responsibility will be used appropriately, and confidentiality and security obligations will be met.
- Ensure that any proposed record linkage is not harmful to individuals or providers and the benefits derived from the record linkage are clearly in the public interest.

DALHOUSIE UNIVERSITY
Health Data Nova Scotia

Data Access Request Instructions

The DAC meets monthly to consider requests. You will be offered the opportunity to attend a meeting to review your application with members of the Committee. If your request for access to data is approved, you will be required to sign appropriate contractual agreements that outline the specific conditions under which the data will be released.

If you have any questions regarding your application, please feel free to contact: hdns@dal.ca

RESEARCH ETHICS BOARD

Please note: Any research project must be reviewed/approved by a research ethics board established and operating in conformity with the Tri-Council Policy Statement "Ethical Conduct for Research Involving Humans" adopted in August 1998 by the Medical Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada and includes any amendments or successor statements. It is the responsibility of the investigator(s) to submit an application to Research Ethics.

PRIVACY

In Nova Scotia, the *Personal Health Information Act (PHIA)* governs the collection, use, disclosure, retention, disposal and destruction of personal health information. *PHIA* recognizes both the right of individuals to protect their personal health information and the need for custodians to collect, use and disclose personal health information to provide, support and manage health care. Data held at HDNS is collected by the DHW without expressed consent. When access to line-level data is requested, the requestor(s) is required to address the impracticability clause in *PHIA,* which requires investigators to seek consent unless doing so is impracticable, which means a degree of difficulty higher than inconvenience or impracticality, but lower than impossibility.

In keeping with the principle that privacy of individuals will be protected as a top priority, requestor(s) are not permitted to release data from cells containing fewer than 5 individuals. If such information is important to the outcomes of the study and the risk of identifying individuals in such cells is low, then it is possible for the investigator to apply for a variance to this rule in a specific case. If this is the case in a specific project, the principal investigator must inform HDNS that they will be requesting permission to publish tables having less than 5 individuals in a cell. The requestor(s) must then submit the specific request in writing together with the rationale for the variance to the HDNS DAC and the DHW. Notification of the decision will be sent to requestor(s). Requestor(s) must also commit to the HDNS policies of pre-submission review of publications to ensure that confidentiality and privacy have been maintained.

COMPLETING AND SUBMITTING THE DATA ACCESS REQUEST

The information you supply on the Data Access Request (DAR) form will be used to evaluate your request for access to HDNS data holdings. Please download the form from the HDNS website, complete all applicable sections of the form and provide any additional information that is requested.

1) Before filling out the form, familiarize yourself with the HDNS datasets and variables that you are applying to access by reviewing the HDNS Data Dictionaries.

2) It is strongly recommended that you discuss with HDNS the feasibility of addressing the questions in your project and the guidelines and limitations for the data in each field that you plan to use.

Data Access Request Instructions

3) List all team members, their affiliation and specific role (e.g., investigator, data analyst, research assistant, statistical or clinical consultant, data collection, etc.) in the proposed project in section1 - Project Team. A rationale must be provided for all team members.

4) Identify the funding source, peer review status and/or purpose of the project in section 2 - Project Details.

5) Provide a detailed description of the project including the objectives, methodology, potential benefits and harms and risks, in section 3 - Project Outline of the DAR Form. Attach the full project proposal as **Appendix 1**.

6) Specify whether the project will involve direct access with potential participants in section 4 - Consent; if so, include a copy of the introductory letter that will be sent to the potential participants as well as the information, questionnaires, and any other materials that potential participants will receive as **Appendix 2**. If participants will be asked to provide informed consent for this project include consent form as **Appendix 3** or describe why consent is not being sought.

7) In section 5 - Data Requested attach a full listing of all variables that you are requesting to be included in the "Project Dataset" as **Appendix 4** using the template provided. Organize these variables by objective and source dataset. Only the minimum dataset in the least identifiable form required to fulfill the project objectives will be considered for data access.

8) List individual diagnostic or procedure codes required for the project in the space provided or provide as **Appendix 5**.

9) For projects seeking to link/match external data with HDNS data, provide a list of external datasets (and custodian), describe the nature of the linkage, and include a description of the linkage procedure that will be used to maintain privacy and confidentiality. If multiple linkages will occur you must include a flow diagram as **Appendix 6**. Proof of custodian approval or request for data access will be required.

10) If you are using external datasets, attach a full listing of all variables that you are requesting to be included in the "Project Dataset" as **Appendix 7** using the template provided. Organize these variables by objective and source dataset. Only the minimum dataset in the least identifiable form required to fulfill the project objectives will be considered for data access.

11) In section 6 - information practices, indicate who will have access to the project dataset, where the remote terminal will be located, what level of data (aggregate vs. person-level) will be accessed, Indicate security measures (e.g., physical, technical and administrative controls and safeguards) that will be in place.

12) Describe how you intend to share and/or publish the results of your project, providing detail on audiences and the format in which data/results will be presented in section 7 - Dissemination of the Project Results.

13) Send the completed Application Form, including the appropriate Appendices 1 through 7, by e-mail to HDNS at hdns@dal.ca.

    It is <u>not</u> necessary to send the *Contractual Agreement for Data Access and Management* or *Confidentiality Agreement* with your completed DAR. You will be asked to complete these agreements, if you receive data access approval.

14) You are encouraged to attend the HDNS DAC meeting at the time of review to answer questions raised during the Committee discussion. If it is not possible for the principal

investigator(s) to attend the meeting, a co-investigator with knowledge about the project may attend.

Once your application has been reviewed and fully approved (i.e., all clarifications and questions have been answered), your data request will be placed into the data request queue. Data requests typically take 4-8 weeks, depending on the complexity of the project. Data requests requiring linkages can take longer. We recommend that investigators begin the data request process as early as possible to limit delays in obtaining data.

DEFINITIONS

Data Linkage

The bringing together of two or more records of personal health information to form a composite record. Linkage allows information on an individual from one dataset to be linked to information on the same individual from another dataset. Using the linked data makes it possible to gain a more comprehensive understanding than could be obtained from either data source individually.

Data Matching

The creation of individual identifying health information by combining individual identifying or non-identifying health information or other information from two or more databases without the consent of the individuals who are the subjects of the information.

Impracticable

A degree of difficulty higher than inconvenience or impracticality but lower than impossibility.

Privacy and Confidentiality

Privacy is the right of individuals to determine when, how and to what extent they share information about themselves with others.

Confidentiality is the obligation of an organization or custodian to protect the information entrusted to it and not misuse or wrongfully disclose it.

Personal Information

Personal Information is recorded information about an identifiable individual, including:

- the individual's name, address or telephone
- the individual's race, national or ethnic origin, color, or religious or political beliefs or associations
- the individual's age, sex, sexual orientation, marital status or family status
- an identifying number, symbol or other particular assigned to the individual
- the individual's fingerprints, blood type or inheritable characteristics
- information about the individual's health care history, including a physical or mental disability
- information about the individual's educational, financial, criminal or employment history,
- anyone else's opinions about the individual, and
- the individual's personal views or opinion, except if they are about someone else.

Personal Health Information

Personal health information is identifying information about an individual, whether living or deceased, and in both recorded and unrecorded forms, if the information:

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

- 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
- relates to payments or eligibility for health care in respect of the individual
- 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
- is the individual's registration information, including the individual's health-card number, or
- identifies an individual's substitute decision-maker.

Types of Information

a) **Identifiable information** – Information that may reasonably be expected to identify an individual, alone or in combination with other available information.

b) **Directly identifying information** – Information that identifies a specific individual through direct identifiers (e.g., name, social insurance number, personal health number).

c) **Indirectly identifying information** – Information that 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 characteristics).

d) **Coded information** – Information that has had direct identifiers removed and replaced with a code. Depending on access to the code, it may be possible to re-identify specific participants.

e) **Anonymized information** – Information that is irrevocably stripped of direct identifiers; no code is kept to prevent future re-linkage, and risk of re-identification of individuals from remaining indirect identifiers is low or very low.

f) **Anonymous information** – Information that never had identifiers associated with it (e.g., anonymous surveys) and risk of identification of individuals is low or very low.

g) **Identifying information** – Information that identifies an individual or, where it is reasonably foreseeable in the circumstances, could be utilized, either alone or with other information, to identify an individual.

h) **De-identified information** – Information that has had all identifiers removed that

(i) identify the individual, or

(ii) where it is reasonably foreseeable in the circumstances, could be utilized, either alone or with other information, to identify the individual.

Levels of Identifiability

a) **Aggregate:** Data that cannot be linked to individuals. For data release purposes, this term generally refers to data that has been grouped with all cells having 5 or more individual patients or subjects. In addition, presentation of data must ensure that summary and other measures cannot allow back-calculations to individual values. Aggregate data for distinct populations or specific communities may still be identifiable or sensitive enough that additional measures to protect privacy and confidentiality may be required prior to release of data tables.

b) **Person-Level (De-identified):** Individual level data that has had all direct and indirect identifiers removed. Direct identifiers are those that are specific to the individual patient or subject and include such variables as name or health card number. Indirect identifiers

are those that could be used alone or in combination to identify an individual patient or subject, such as birth date or full address. For data release purposes, individual data are "coded" as per the definition above. That is, direct identifiers are removed from each patient/subject record in the analysis file and replaced with a unique identifier that is specific to the approved project. Indirect identifiers are modified to reduce the risk of re-identification. Examples include replacing birth date with age and replacing a full postal code with the forward sortation area (first 3 digits) of the postal code.

c) **Person-Identifiable:** Individual level data that contains direct or indirect identifiers. For data release purposes, patient/subject consent is usually required. Occasionally an indirect identifier is required to complete a project. The researcher must provide a strong argument for including an indirect identifier in the analysis file. *Note: For any linkage project, direct identifiers will be required to perform the linkage. These identifiers will be used to link the files required for the approved project, then the file will be coded as described above under "Person-Level (De-identified)." The data in the analysis file is then considered person-level (de-identified).*

d) **Health Care Provider-Identifiable:** Individual care providers can be identified by name, practice number, specialty, practice parameters, or location of practice (e.g., cardiac surgeon practicing in XX County when there is only one cardiac surgeon in that county). For data release purposes, caregiver consent is usually required. The researcher must provide a strong argument for including health care provider-identifiable data in the analysis file.