DALHOUSIE UNIVERSITY

HDNS Guidelines for Data Linkages

Data Linkages

Linkages are conducted at HDNS between datasets held by HDNS ("internal linkages") and between datasets held by HDNS with datasets held externally ("external linkages"). All linkages are performed by HDNS and linked data reside on the HDNS Secure Data Platform (SDP). The processes for each are as follow:

Internal Linkages:

- HDNS extracts the study datasets for a specified study population from HDNS data holdings, including encrypted common identifiers (eIDs) such as patient health card numbers (HCNs) or provider identification numbers.

- HDNS links datasets using eIDs to create a project dataset.

- HDNS removes the eIDs and replaces with a randomly generated study ID.

- All analyses are conducted on the HDNS SDP either by HDNS analysts or accredited research analysts. External analysts will be provided with project based access to the HDNS SDP and the project dataset. The HDNS SDP can be accessed either remotely or at a secure HDNS workstation. Line-level data cannot leave the HDNS system without approval.

- Access for analyses, either by the HDNS, or external analysts is only granted once all data access approvals are in place and agreements are signed.

External Linkages:

External datasets are data that are not currently part of HDNS data holdings, for example: (i) data from a clinical trial, research study or survey; (ii) data from patient registries including disease or population–based data; or (iii) data from other organizations or custodians.

For external data linkages HDNS:

- does not accept external datasets that contain direct identifiers such as names or street addresses.

- only accepts common identifiers such as HCNs, provider numbers and social insurance numbers (SINs) that have been encrypted by through the approved encryption process to create eIDs.

- only accepts dataset that have approval from the custodian of the external dataset.

Process for external data linkage:

- If the external dataset has direct identifiers they must be removed, only the common identifier (e.g. HCNs) will be used to link to HDNS datasets.

- The project team or external data custodian sends a file with only a list of common identifiers (HCNs) along with a corresponding study ID number to the Nova Scotia MSI vendor, Medavie without including any other variables. Medavie will only accept these two

variables. Medavie scrambles the common identifier using an encryption algorithm and sends the encrypted HCN and the corresponding study ID to HDNS.

- If the project team has additional approved variables to send to HDNS, they create a data file of these variables that includes the study IDs, but not any common identifiers. This file is then sent directly to HDNS using an approved secure transfer method.

- HDNS links the researcher's data file to the project dataset using the study ID.

- For details on approved methods of sending data to Medavie for encryption and to HDNS for data linkage, please contact HDNS (hdns@dal.ca).

- All analyses are conducted on the HDNS SDP either by HDNS analysts or accredited research analysts. External analysts will be provided with project based access to the HDNS SDP and the project dataset. The HDNS SDP can be accessed either remotely or at a secure HDNS workstation. Line-level data cannot leave the HDNS system without approval.

- Access for analyses, either by the HDNS, or external analysts is only granted once all approvals are in place and agreements are signed.

Figure 1: External Data Linkages

This figure is intended to illustrate the process for linking, when needed. For most studies, multiple analysis datasets are created, depending on the types and structure of the data required.
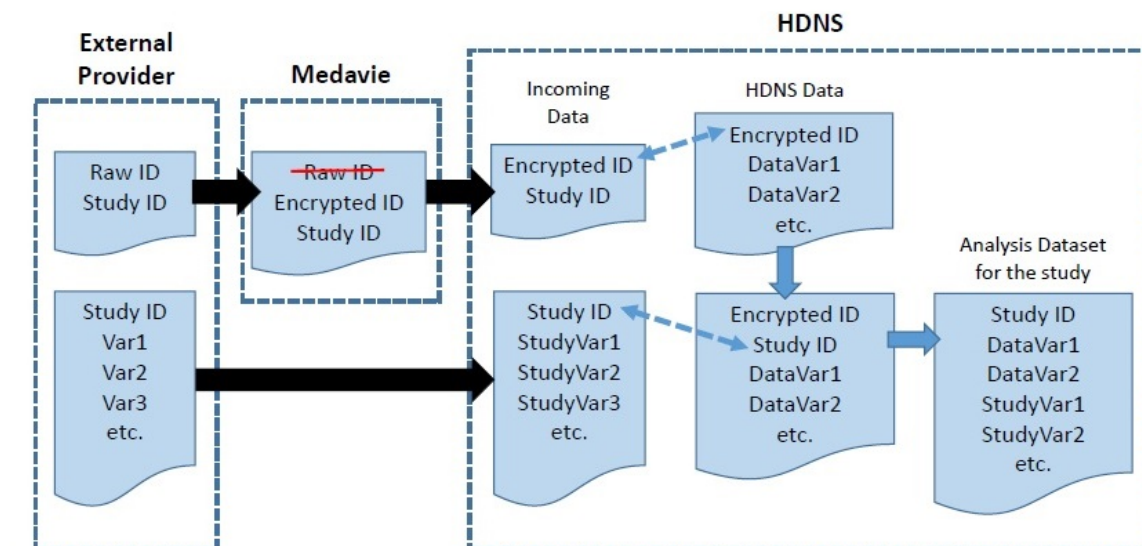
LEGEND

Secure Delivery

Linkage

Resulting Dataset

DEFINITIONS
External Provider: Any data provider other than HDNS (e.g., study team, RCP, CCNS)
Raw ID: Any unique individual identifier that can be used to link to HDNS data (e.g., patient health card number, provider medical board number)
Encrypted ID: The "raw ID" encrypted per Medavie protocol.
Study ID: A unique individual identifier assigned by the study team as a proxy for identifying data such as patient health card number.