



Manitoba Center for Health Policy

Data Request Guidelines for Project Specific Data

Dave Towns
Revised May 26, 2022

- Manitoba Health has implemented a secure file transfer system for data delivery. See pages 2 and 3 for details about requesting a secure file transfer user account.

Table of Contents

Introduction.....	2
Providing data to Manitoba Health	2
Procedure for transferring files to Manitoba Health	3
What data should be included or excluded in your data request?	3
Approaches for describing the requested data	5
Data modifications at Manitoba Health.....	8
De-identification of data at Manitoba Health	8
When the data arrives at MCHP	9
What about additional data requests to fix data issues?	9
Questions to consider before sending your data request	9

Introduction

This document is intended to provide guidance on developing a data request that outlines your project specific data requirements and de-identification needs.

All data maintained at the Manitoba Centre for Health Policy (MCHP) must be de-identified. Manitoba Health will receive data that you request from a source agency and transform it into a format that can be used at MCHP. This document will describe in detail the de-identification process.

Providing data to Manitoba Health

Before requesting any data be sent to Manitoba Health for de-identification, confirm that all the appropriate approvals are in place with the MCHP Repository Access Coordinator.

This might include paperwork to meet MCHP requirements; Provincial Health Research Privacy Committee (PHRPC) requirements; University of Manitoba (UM) ethics requirements (e.g., HREB); requirements associated with any non-Health data providers; and UM-MCHP Researcher Agreement requirements.

Contact details for Manitoba Health and MCHP are provided in the sample data request example in Appendix A.

Manitoba Health and MCHP can work with a variety of data file formats (e.g., Excel, Access, SPSS, SAS datasets, comma-separated values (CSV) file with double-quoted textual content, structured text files, etc.). For other types of application data, it may be appropriate to export the data into a more standard format.

Some applications can provide additional password security that restricts access to the information stored inside a data file (e.g., Excel, SPSS). These passwords should not be used since they can make it difficult to import the data.

The delivery of source agency data and other documentation must be done securely. For most projects this will be accomplished by:

- Placing all data files in one or more password encrypted archives. The password encrypted archives will then be delivered in the following ways:
 - Manitoba Health will electronically receive your archives (a secure file transfer account is required, see the next section for details).
 - MCHP will receive your archives on CD-ROM, DVD-R or USB media. It should be noted that DVD media may have a maximum file size limit of 2 GB for files that are burnt onto a DVD. Larger sized files may not be readable from the DVD. USB media can also have a similar file size limit of 4 GB. You may need to use multiple archives below these file size limits to address storage medium limitation.

- Canada Post or email should not be used to deliver confidential data. MCHP can receive your non-identity secured data, on appropriate media, by courier or by hand delivery to our department reception area.
- The password for the secured data archive will be sent by email to the appropriate individual.

Procedure for transferring files to Manitoba Health

Files containing personal health information that are required to be disclosed to Manitoba Health (MH), must be sent using MH's secure file transfer system.

When a file is ready to send to MH, please contact MH to have an account set up:

Provincial Information Management and Analytics (PIMA)
 Manitoba Health
 4th Floor – 300 Carlton Street
 Winnipeg, MB
 R3B 3M9
 Tel: 204-786-7139
 Email: pima@gov.mb.ca

MH will set up an account using your email address. You will be notified through email once it has been created and will be provided with a username and temporary password.

All files sent using the secure file transfer must be password protected and the password is to be sent to the recipient separately either by email or by telephone.

What data should be included or excluded in your data request?

When you request data from the source agency, their data collection system will often be organized into multiple tables of very specific types of content. This might include a client related table with personal and demographic details; transactional tables like appointments, visit details, medications given, tests requested, billing records, etc.; and sometimes tables that define the codes used within the various system tables and their definition (e.g., M = Male, F = Female; 1 = Yes, 0 = No).

Your project data request might be for a single consolidated table that is based on related data elements from several tables. In this situation, you are identifying just the data elements you need and excluding other content that has no value to your project.

For more complex data systems, you may need to request a group of related tables that will be joined during analysis at MCHP. In such a scenario, please request a data model that outlines the primary key used to join these tables together correctly. You might request that the table data be filtered to only provide a subset of the data that you are interested in for your project analysis. You may also decide that only specific data elements in each table should be provided, thus reducing the amount of data being transfer to Manitoba Health and MCHP.

Manitoba Health will be interested in client personal details to establish a unique scrambled PHIN that will be used for data analysis at MCHP. The following is a summary of data elements that are used in de-identification and record linkage processes and may be contained in client related tables:

- a unique Client Identifier (key) that is permanent, unique to the individual, and not based on their name or other personal identifiers
- Surname, Given Names
- Sex (and how it is coded)
- Date of Birth (and how it is formatted) – Age might work but is less helpful
- Address information and/or Postal Code
- Manitoba Health ID Number
- Personal Health Identification Number (PHIN – 9-digit number)
- Registration Number (REGNO – 6-digit number)
- The date when data was captured. This date can help with synchronizing client specific details like age, postal code, surname, and REGNO with the Manitoba Health registries

Please also ask for documentation on any known data related issues like:

- Data coding changes and issues over time
- Data elements that have been poorly filled in
- Periods in the data that may have been impacted by policy or process changes
- Any other types of known data quality issues

This type of documentation will help Manitoba Health with record linkage and with understanding potential problems within the data that you receive for your project.

Of note, both Manitoba Health and MCHP should not be provided with a client's:

- Social Insurance Number

Additionally, MCHP should not be provided with the following data elements:

- Surname, Given Names
- Address
- original unscrambled version of the Personal Health Identification Number
- Telephone, FAX numbers
- e-mail address
- Some textual data elements that provide comments or notes that might disclose personal details (names, addresses, phone numbers, etc)

Note: the original Registration Number (REGNO), Sex, Date of Birth, and Postal Code may be provided to MCHP. These data elements may be needed to document unlinked responses and validate the PHIN's returned by Manitoba Health.

Additional types of documentation may be useful to ask for. The following are examples of documentation you may want to request from the source agency:

- Any data collection forms that may be used to gather data manually before it is entered into the data collection system
- A copy of survey documents used to collect client data in electronic format
- Documentation on the data codes used to represent descriptive textual data values. Often with computerized systems, data is selected from a pull-down list of values. The selected value may be stored in the data table as a coded value (numbers and/or letters) rather than the actual list value.
- As mentioned previously, data codes and their descriptive textual data values may be stored in a table that is used by the data collection system. This table should be requested if it is available.

Approaches for describing the requested data

In Appendix A, a simple data request has been provided for a non-health data trustee. It outlines the data being requested and contact details for Manitoba Health and MCHP.

Depending on the complexity of your data request, you may need to outline the details of exactly what data is being requested. If your request details are short, they might be included directly in the data request letter. More detailed requests might include an appendix to provide content like the following:

- a list of tables that are being requested
- a lists of data elements for each table
- details about data elements that are not needed and should be removed
- details about data elements that need to be modified (e.g., create a flag indicating a yes/no response)
- any data modifications that may be required at Manitoba Health

The following is a simple example showing how you might request participant details:

The following data elements have been identified for this project:

- personal details associated with program participant (child)
 - first name, middle name or initial, last name
 - birthdate
 - gender
 - street address, postal code
- If available, personal details (name/address details) associated with the parent or guardian of the program participant that would aid in identifying the child at Manitoba Health.

In other situations, it may be appropriate to identify all the data tables that you will need from a specific database system:

For this project, please include all information available for dates beginning on July 11, 1988, up to the last date or present, including:

- tables:
 - Registry
 - Laboratory Test
 - Pharmacy
 - Tissue

Any documentation describing each of these extracted database tables.

For very detailed requests it may be useful to outline exactly what you are asking for in an appendix.

An appendix would contain the data request details. In the following example, hundreds of data tables were identified. Some were not wanted. Some were completely requested while for others, only a portion of the tables were requested. In this example of non-health data, most tables were sent directly to MCHP (e.g., they contain no client identifiable data elements). A single table was sent to Manitoba Health with confidential client details that were provided to create a crosswalk/linkage table used at MCHP:

<p>Changes to this document since the previous extraction have been identified with a star: ★</p> <p>The following tables have been added or updated:</p> <table> <tr> <th></th> <th></th> <th>Page Number</th> </tr> <tr> <td>1</td> <td>tbl_district</td> <td>17</td> </tr> <tr> <td>2</td> <td>taofferdecisionreason</td> <td>31</td> </tr> <tr> <td>3</td> <td>tblclientincome</td> <td>57</td> </tr> </table>					Page Number	1	tbl_district	17	2	taofferdecisionreason	31	3	tblclientincome	57
		Page Number												
1	tbl_district	17												
2	taofferdecisionreason	31												
3	tblclientincome	57												
<p>LIST OF TABLES THAT ARE NOT REQUIRED</p> <ol style="list-style-type: none"> adminentitybatch adminentitybatchnumber apdr_household_type apdr_income_type apdr_unit_type building_rp ccra_summary chargeback company csclient_rp csclientbanking csclientname csclientsin csvalformletter current_balances current_differences dba_table_stats dba_yesterday_tables dropped_usertables glaccount glaccountchangelog glaccountcode glaccountmask glaccountstatus glacct_mask glcondition gldistrict gldsinterface gldsinterface_initial glmaskingdata glminterface glminterface_initial gloperand 														
5/7/2015	Tenant Management System - Table Request	Page 1 of 64												

Order	Column Name	Type	Request for Value Added Fields
In the tables that follow, clientid has been highlighted in yellow to draw attention to this data field. The clientid field should be scrambled to eliminate the potential to identify a client in the TMS system.			
Example:			
1	clientid	integer	SCRAMBLE
LIST OF TABLES - ALL COLUMNS REQUIRED			
adminentitytype			
Order	Column Name	Type	
1	code	integer	
2	abbreviation	varchar	
3	description	varchar	
adminfunctiontype			
Order	Column Name	Type	
1	code	integer	
2	abbreviation	varchar	
3	description	varchar	
areaunit			
Order	Column Name	Type	
1	code	integer	
2	abbreviation	varchar	
3	description	varchar	
bldgbalconytype			
Order	Column Name	Type	
1	code	integer	
2	abbreviation	varchar	
3	description	varchar	
5/7/2015	Tenant Management System - Table Request	Page 9 of 64	

Order	Column Name	Type	Request for Value Added Fields
csclient			
Order	Column Name	Type	
1	clientid	integer	
2	namefirst	varchar	
3	namelast	varchar	
4	namemiddle	varchar	
5	birthdate	date	
6	deceaseddate	date	
7	sex	char	
8	mhszno	varchar	
9	divlcno	varchar	BOOLEAN INDICATOR - PRESENCE OF VALUE
10	vehlcno	varchar	BOOLEAN INDICATOR - PRESENCE OF VALUE
11	sin	varchar	BOOLEAN INDICATOR - PRESENCE OF VALUE
12	citizenshipstatus	integer	
13	languagespoken	integer	
14	datecanres	date	
15	immigranticategory	integer	
16	areacode	char	
17	homephone	varchar	
18	workareacode	char	
19	workphone	varchar	
20	altareacode	char	
21	aliphone	varchar	
22	aliphonemote	varchar	
23	fasareacode	char	
24	fax	varchar	
25	email	varchar	BOOLEAN INDICATOR - PRESENCE OF VALUE
26	landlordid	integer	
27	username	varchar	
28	createdate	date	
29	changedate	date	
30	latestapplication	integer	
csclientfreeaddress			
Order	Column Name	Type	
1	clientid	integer	SCRAMBLE
2	residence1	varchar	
3	residence2	varchar	
4	residence3	varchar	

Some data elements were modified into a new format

GRAY data elements were not wanted in this table

Order	Column Name	Type	Request for Value Added Fields
LIST OF TABLES - TO MANITOBA HEALTH			
Combined Table Content: csclient + csclientfreeaddress			
Order	Column Name	Type	Associated Table for Field Content
1	clientid (SCRAMBLE)	integer	csclient
2	namefirst	varchar	csclient
3	namelast	varchar	csclient
4	namemiddle	varchar	csclient
5	birthdate	date	csclient
6	deceaseddate	date	csclient
7	sex	char	csclient
8	mhszno	varchar	csclient
9	residence1	varchar	csclientfreeaddress
10	residence2	varchar	csclientfreeaddress
11	residence3	varchar	csclientfreeaddress
12	residencepostalcode	varchar	csclientfreeaddress

The above table combines the client related details (name, sex, birthdate, any medical #'s) with the associated residence address of the client. These details will be used to generate a crosswalk file for MCHP.

Data modifications at Manitoba Health

Some of the data elements that you are requesting may need to be modified by Manitoba Health. Examples of data modifications might include:

- changing a physician's name into their associated MD number
- provide a yes/no flag to identify that a data element had identity related data available. Examples of this might include: a social worker flag instead of the social worker's name to identify that a social worker was assigned to the client. Other examples might include indicators for a driver's license or personal email.
- large textual fields might contain useful details that may help with your analysis. But they may also contain confidential information that should be cleansed before the data arrives at MCHP (e.g., remove names, phone numbers, addresses, etc.)

De-identification of data at Manitoba Health

The processes of transferring data to MCHP requires de-identification. Manitoba Health will take demographic/sensitive data elements and remove them from the requested database. A scrambled Personal Health Identification Number (PHIN) is added to the database to permit linkage of your project data to other databases in the MCHP repository.

Appendix B provides a detailed overview of the two approaches for transferring data to Manitoba Health for de-identification depending on the source of the data. The following is a short description of these two processes:

- Manitoba Health Data Trustee
 - i. Manitoba Health removes PHIN, names, addresses, and other sensitive data elements and adds scrambled PHIN, sending the whole data system to MCHP.
- Non-Manitoba Health Data Trustee
 - i. The source agency/trustee sends a demographic subset (names, addresses, effective date of address, date of birth, scrambled person ID, and sometimes PHIN) to Manitoba Health.
 - ii. The source agency/trustee at the same time sends the entire database, minus the identifying fields in the demographic subset to MCHP. The scrambled person ID is included in the demographic subset sent to MCHP. MCHP will prepare the received project data, while awaiting the delivery of a crosswalk/linkage file from Manitoba Health.
 - iii. Manitoba Health adds scrambled PHIN to the demographic subset, removes all identifying information, and sends a crosswalk/linkage file (scrambled PHIN, scrambled person ID, and linkage results) to MCHP.
 - iv. MCHP stores the files, unlinked. On an approved project-by-project basis, the crosswalk/linkage file is linked to the main source agency data files.

Appendix B will be useful for any discussions you may have with the source agency / trustee about how their data will be de-identified.

When the data arrives at MCHP

Data that arrives at MCHP will consist of one or more data files in varying formats. Incoming files are converted to a form suitable for use in the data analysis environment and are made internally coherent. Some preliminary testing, validation, data cleaning and recoding of values to conform to MCHP standards takes place. Format files (codes and labels), if available, may be converted into SAS.

At this stage, linkage among the data files comprising the database and linkage to other resources in the MCHP repository does not normally take place.

Once a crosswalk/linkage file is available, the Internal Reference IDs from the crosswalk will be checked against the previously received data file IDs. The encrypted PHIN from the crosswalk will be evaluated to determine how well it links to the MCHP repository.

Data access privileges will be setup, so that the data analyst(s) assigned to your project can begin working with the project data.

What about additional data requests to fix data issues?

Once the data analysts start working with the data, it is possible that data issues may be identified. This may require that you request the source agency to re-extract a portion of the data to fix the identified problems. Make sure that Manitoba Health and MCHP are aware of your request by including them in the re-extract request email. These re-extracted files should be handled using the same secure process as the original data.

Questions to consider before sending your data request

1. Who is the data trustee (as defined in PHIA)?
2. Do you have the data Trustee's approvals to disclose and use the requested data in your project? These approvals must be in place prior to any transfer or use of data.
3. Do you need to include a source agency privacy officer on your data request?
4. Also, please make sure you have completed all approval related paperwork associated with using MCHP repository related data as well. For further details, please see the [Applying for access and use](#) section of MCHP's website.

Appendix A – Database / Table Request Examples

In this example you will notice <YELLOW HIGHLIGHTED> content which identifies places where you would enter your project request details.

July 18, 2016

TO: SOURCE_AGENCY_CONTACT,
SOURCE_AGENCY_CONTACT_ROLE
SOURCE_AGENCY_CONTACT_ORGANIZATION_NAME

FROM: YOUR_NAME
ORGANIZATION

RE: Request for extraction of PROJECT_NAME database.

The following data request is for the extraction of the PROJECT_NAME database. This data will be used in the FORMAL_DESCRIPTION_OF_YOUR_PROJECT of research.

PROVIDE DETAILS ABOUT THE DATA YOU ARE REQUESTING. THIS MAY BE A LIST OF TABLES, A LIST OF DATA ELEMENTS, OR AN ATTACHMENT IF THE DATA IS COMPLEX.

Please add any additional system keys that may be associated with linking any underlying tables being requested. Feel free to add additional data elements that might be of interest for these tables that may have been missed in the above list.

We would appreciate receiving documentation that describes the PROJECT_NAME database tables, any data element codes/definitions, and technical table specifications that might help with understand the PROJECT_NAME tables.

All the PROJECT_NAME extracted data will be sent directly to Manitoba Health for de-identification.

Shipping details:

To be transferred to Information Management and Analytics at Manitoba Health:

Table(s) containing the contents of ‘Identifying Information’ related data (as outlined in this request, in an attached document or Appendix if needed):

One or more tables will be electronically transferred in a password encrypted archive (WinZip compatible). The archive filename should include enough detail to clearly identify the contents (e.g., “DATABASE-PROJECT_NAME-PHRPC20182019-99-DATE”).

The password(s) are to be sent separately by e-mail to the Provincial Information Management and Analytics (PIMA) department (pima@gov.mb.ca) at Manitoba Health.

Manitoba Health personnel will locate a scrambled PHIN for use at MCHP and will remove all confidential fields from each table.

The remaining fields (including the scrambled PHIN) and all supporting files will then be delivered by secure means to Dave Towns at MCHP.

Contact Information:

Manitoba Health:

Provincial Information Management and Analytics (PIMA)
Manitoba Health
4th Floor - 300 Carlton Street
Winnipeg, MB
R3B 3M9
Tel: 204-786-7139
Email: pima@gov.mb.ca

Files containing personal health information that are required to be disclosed to Manitoba Health (MH), must be sent using MH's secure file transfer system.

Contact PIMA to have an account set up when your secure archive(s) are ready to send to MH. A secure transfer account will be set up using your email address. You will be notified through email once it has been created and will be provided with a username and temporary password.

To be sent to Manitoba Centre for Health Policy:

All other tables that do not contain ‘identifying information’:

The tables will be copied to a password encrypted archive on a CD-ROM, DVD-R or USB media. Please refer to “**DATABASE / PROJECT_NAME / PHRPC2018/2019-99 / DATE**” on the media labels and the filenames. Please note all archived data should be WinZip compatible.

Reference tables and documentation may be shipped in electronic form on the same media as the “**DATABASE / PROJECT_NAME / PHRPC2018/2019-99 / DATE**” extract databases.

The password(s) are to be sent separately by e-mail to Dave Towns at MCHP (dave_towns@cpe.umanitoba.ca).

Contact Information:

University of Manitoba / MCHP:

Dave Towns
Senior Data Management Analyst
Manitoba Centre for Health Policy
Max Rady College of Medicine
Rady Faculty of Health Sciences
University of Manitoba
408-727 McDermot Avenue
Winnipeg, MB R3E 3P5
Ph: (204) 318-5272
dave_towns@cpe.umanitoba.ca

If you have any questions or concerns about this request, please contact **YOUR NAME (PHONE# or YOUR_EMAIL)**

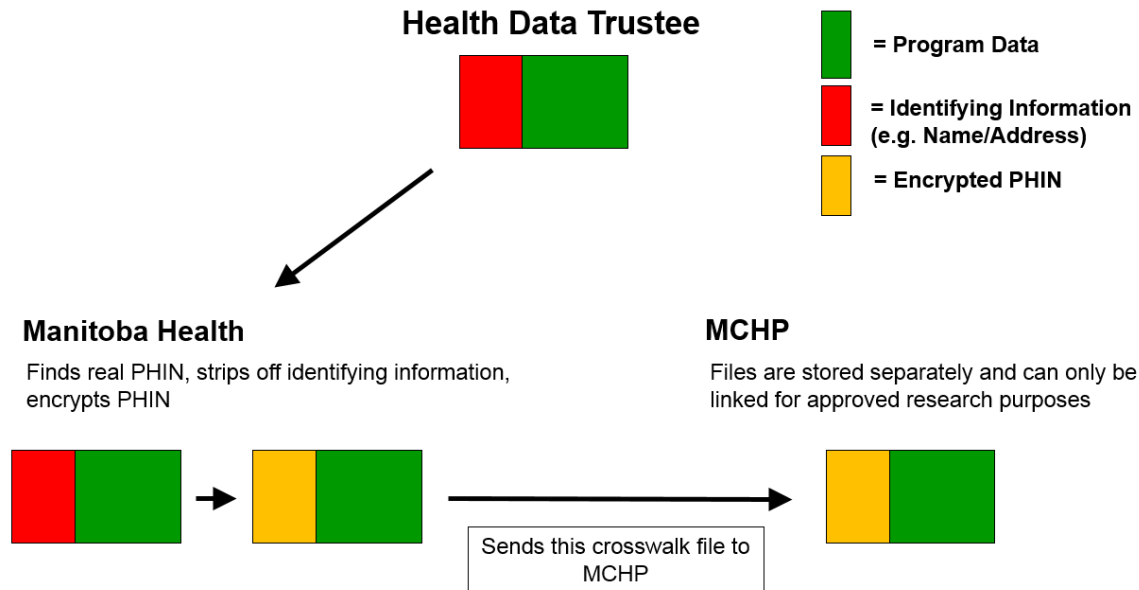
Sincerely,

Your Name
Organization
Address Details
Phone
Email

cc: **OTHER SOURCE AGENCY CONTACTS**
SOURCE AGENCY PRIVACY OFFICER CONTACT
Charles Burchill
Dave Towns

Appendix B – Health Data Trustee

The transfer of data from <SOURCE AGENCY> to the MCHP Repository takes place in several stages. Data privacy is a major consideration. Only anonymized (“de-identified”) data are housed within the MCHP Repository. A multi-stage process involving Information Management and Analytics of Manitoba Health, MCHP and <SOURCE AGENCY> results in a modified version of the data that includes an encrypted PHIN which enables linkages of the anonymized <DATABASE NAME> database with the anonymized MCHP Repository population registry.



Examples of ‘Identifying Information’ might include names, addresses, effective date of address, date of birth, scrambled person ID, and sometimes a Personal Health Identification Number (PHIN). Social Insurance Numbers should not be provided to either MCHP or Manitoba Health.

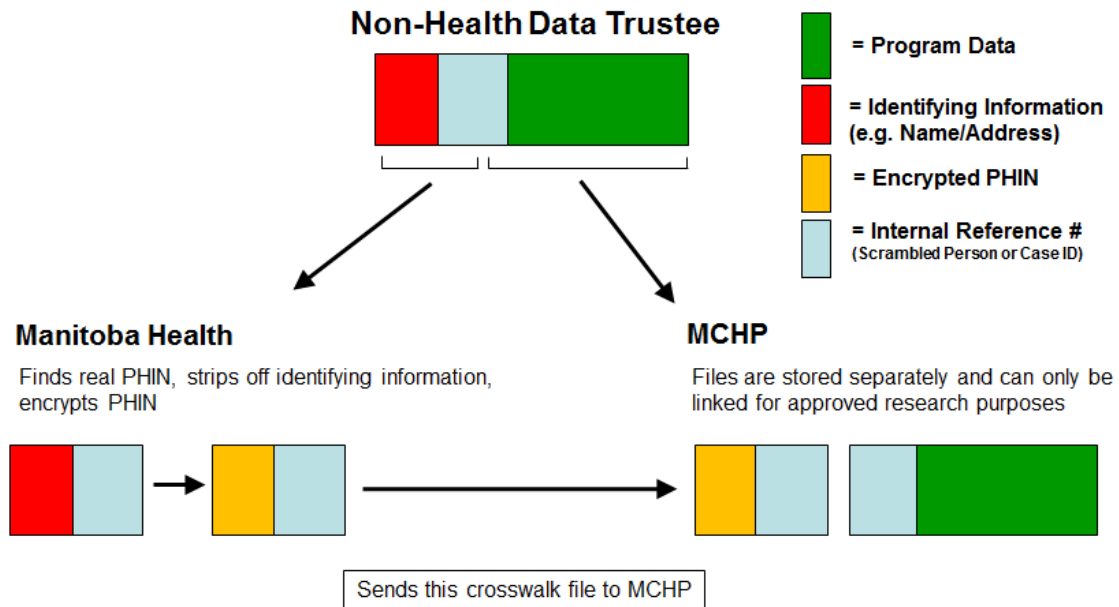
Often with related data tables, a unique ID (key) in your data is used to identify each data record and this ID may also be included in other tables to relate associated data records between tables. If there is any potential concern for this ID being used to identify an individual, it would be warranted to encrypt the ID value, replacing the ID with a new ‘Encrypted_ID’ value.

If the unique ID (key) is always associated with a single person, then the approach/code used for creating the ‘Encrypted_ID’ based on this unique ID (key) should be maintained for future data extractions. Using the same coding scheme for the ‘Encrypted_ID’ will allow for research on individuals who may exist over time in consecutive years of data extractions.

All data and any additional support documents will be provided to Manitoba Health.

Appendix B – Non-Health Data Trustee

The transfer of data from **<SOURCE AGENCY>** to the MCHP Repository takes place in several stages. Data privacy is a major consideration. Only anonymized (“de-identified”) data are housed within the MCHP Repository. A multi-stage process involving Information Management and Analytics of Manitoba Health, MCHP and **<SOURCE AGENCY>** results in the creation of a “cross-walk file” which enables linkages of the anonymized **<DATABASE NAME>** database with the anonymized MCHP Repository population registry.



Examples of ‘Identifying Information’ might include names, addresses, effective date of address, date of birth, scrambled person ID, and sometimes a Personal Health Identification Number (PHIN). Social Insurance Numbers should not be provided to either MCHP or Manitoba Health.

Also note the importance of providing an ‘Internal Reference #’ that is based on a unique ID (key) in your data to uniquely identify each data record. The ‘Internal Reference #’ will replace this unique ID (key) across all other tables that also include the unique ID (key). The ‘Internal Reference #’ will allow MCHP to re-link the Encrypted PHIN from Manitoba Health to the original ‘Program Data’ and link to other associated tables.

If the unique ID (key) is always associated with a single person, then the approach/code used for creating the ‘Internal Reference #’ based on the unique ID (key) should be maintained for future data extractions. Using the same coding scheme of the ‘Internal Reference #’ will allow for research on individuals who may exist over time in consecutive data extractions.

Data tables containing ‘Identifying Information’ are split into two parts, one for Manitoba Health and the second part for MCHP. Any other data tables which do not have any ‘Identifying Information’ can be directly provided to MCHP.