



populationdata^{BC}

A PLATFORM SUPPORTING RESEARCH ON
HUMAN HEALTH, WELL-BEING AND DEVELOPMENT.

POPULATION DATA BC

“Just Give Me My Data”

Researcher Access to Privacy Sensitive Data

Jim Mintha

Population Data BC
University of British Columbia

April, 2018



POPULATION DATA BC SERVICES



**DATA ACCESS
FOR RESEARCH**



**DATA
LINKAGE**



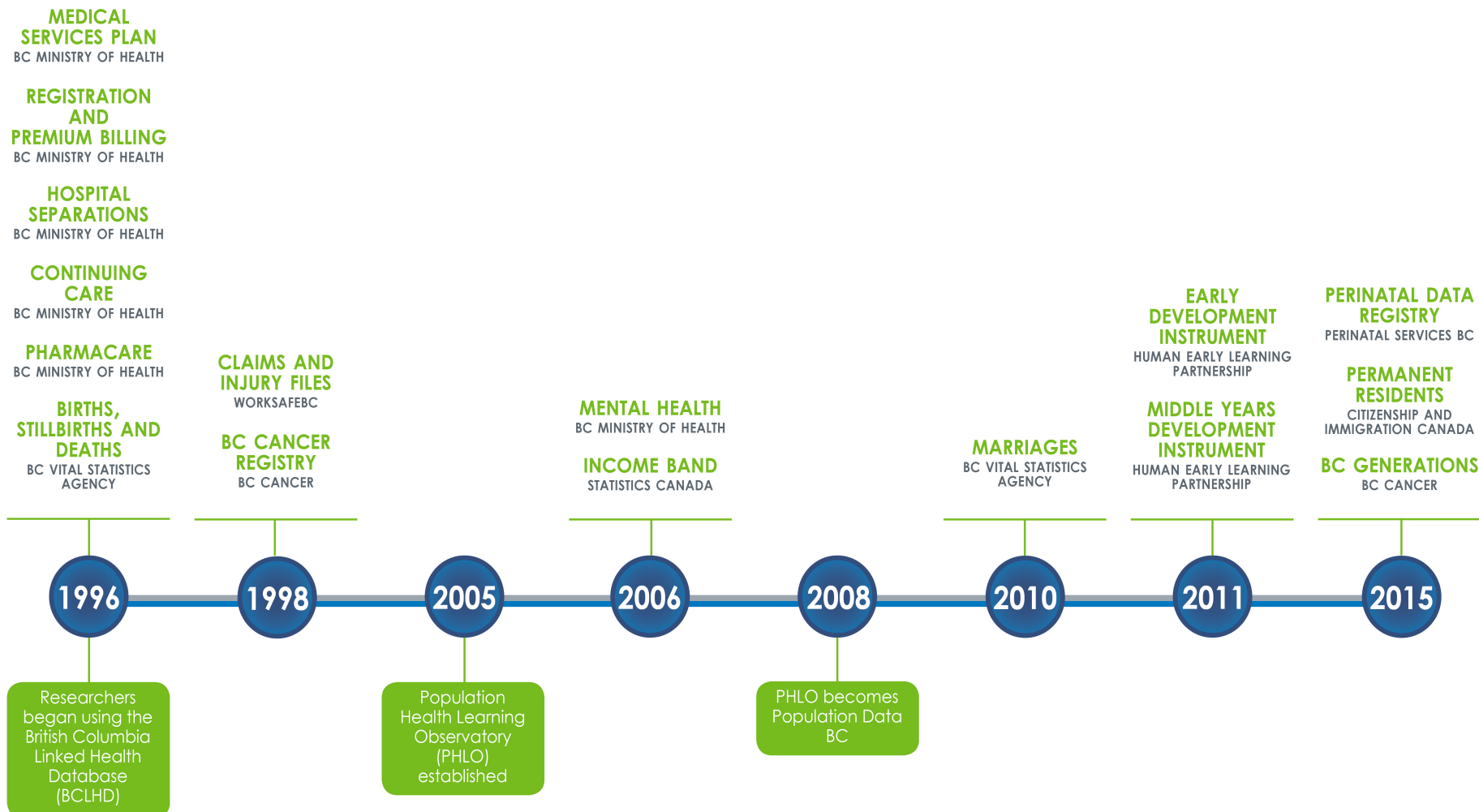
**SECURE DATA
STORAGE AND
ARCHIVING**



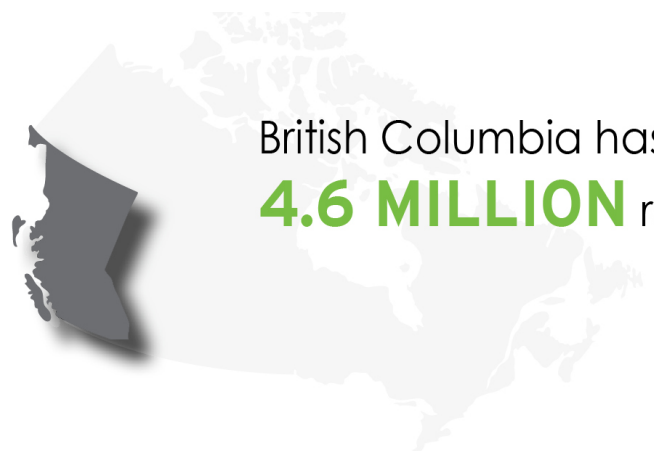
**EDUCATION
AND TRAINING**

POPULATION DATA BC

A brief history



DATA HOLDINGS & POPULATION COVERAGE



British Columbia has
4.6 MILLION residents

We hold data on **ALL** of them
100% coverage

For some people we have

A LOT OF DATA
for other people
VERY LITTLE

HOW MUCH DATA
we have on a person
depends upon their
LIFE EXPERIENCE

Data holdings relate to, for example:

Hospital stays



Education/early
childhood



Visits to a Doctor



Medications



Births, marriages &
deaths



Citizenship



Pregnancy &
childbirth



Disease registries



Workplace injuries



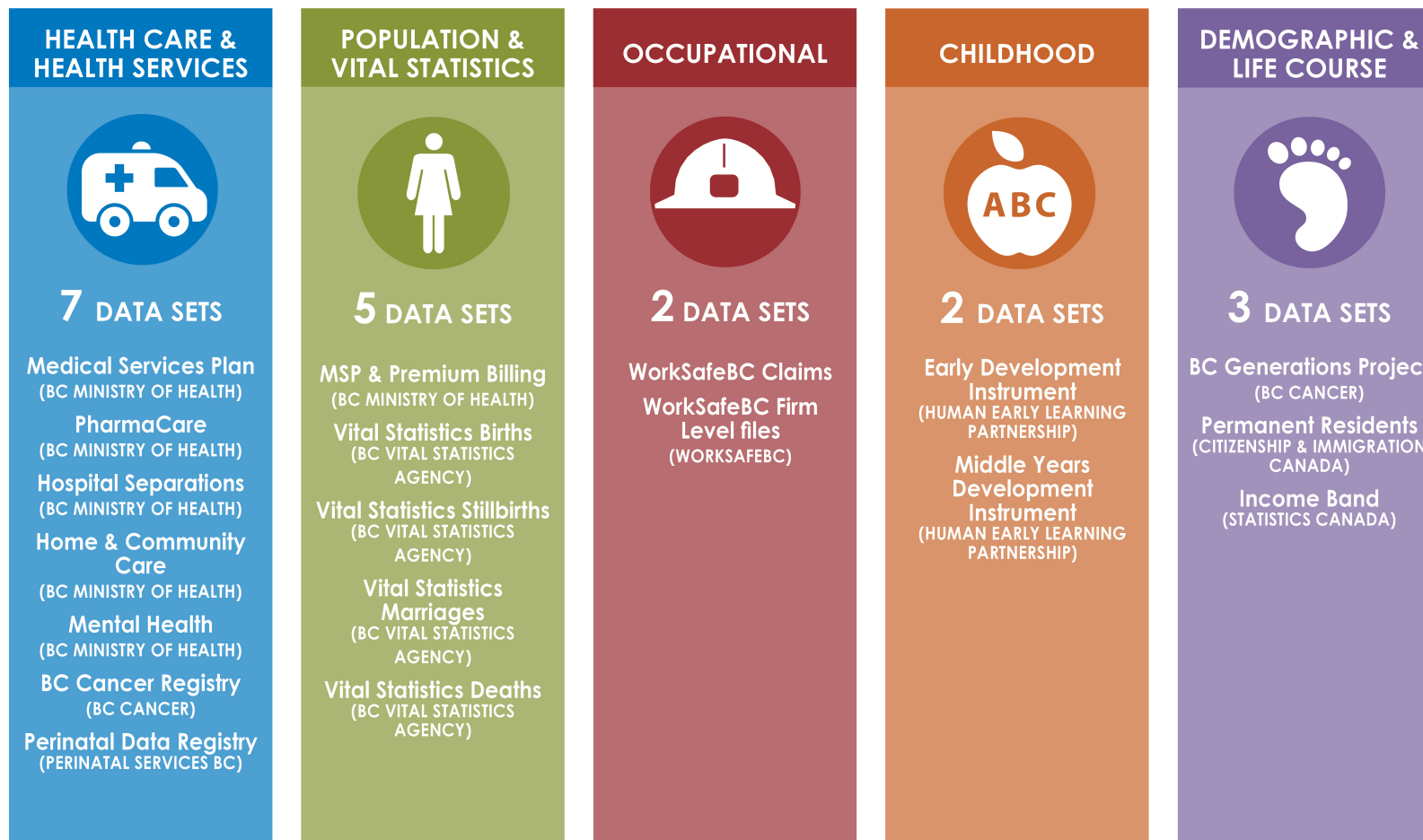
ER visits



DATA SOURCES

Internal data holdings for linkage

19 datasets from 2 Federal and 6 Provincial sources



Legislative Permission:

- **Freedom of Information and Protection of Privacy Act (FOIPPA)** in British Columbia, Section 35 says:

“A public body may disclose personal information in its custody or under its control for a research purpose, including statistical research...”

SECURE DATA STORAGE and ARCHIVING

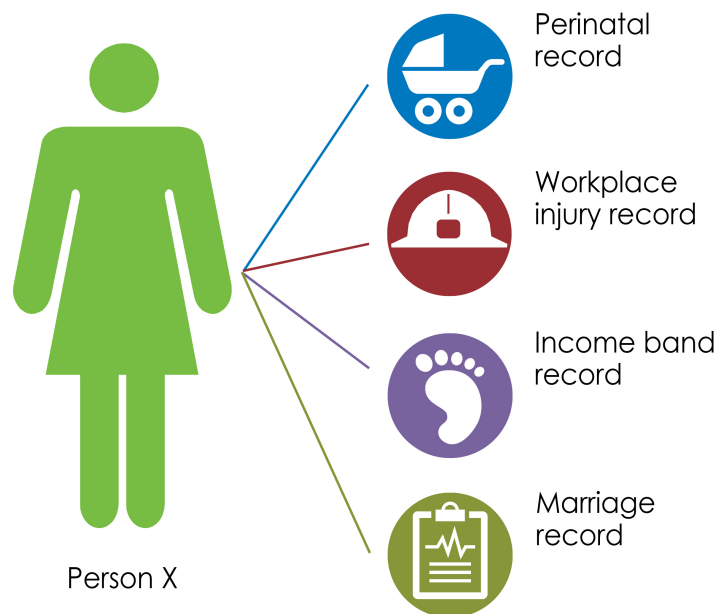
Data protection and privacy

Our mandate: to meet and exceed all relevant legal, ethical and legislative guidelines for data protection and privacy

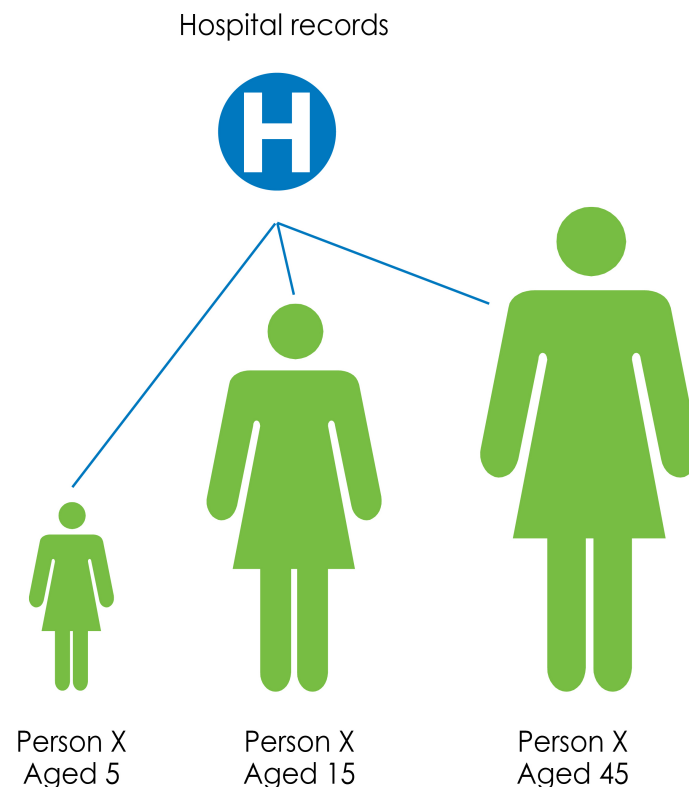
The tenets of individual privacy and data protection are embedded in all aspects of operations, and include externally audited:

- Physical measures
- Technical measures
- Procedural measures

DATA LINKAGE

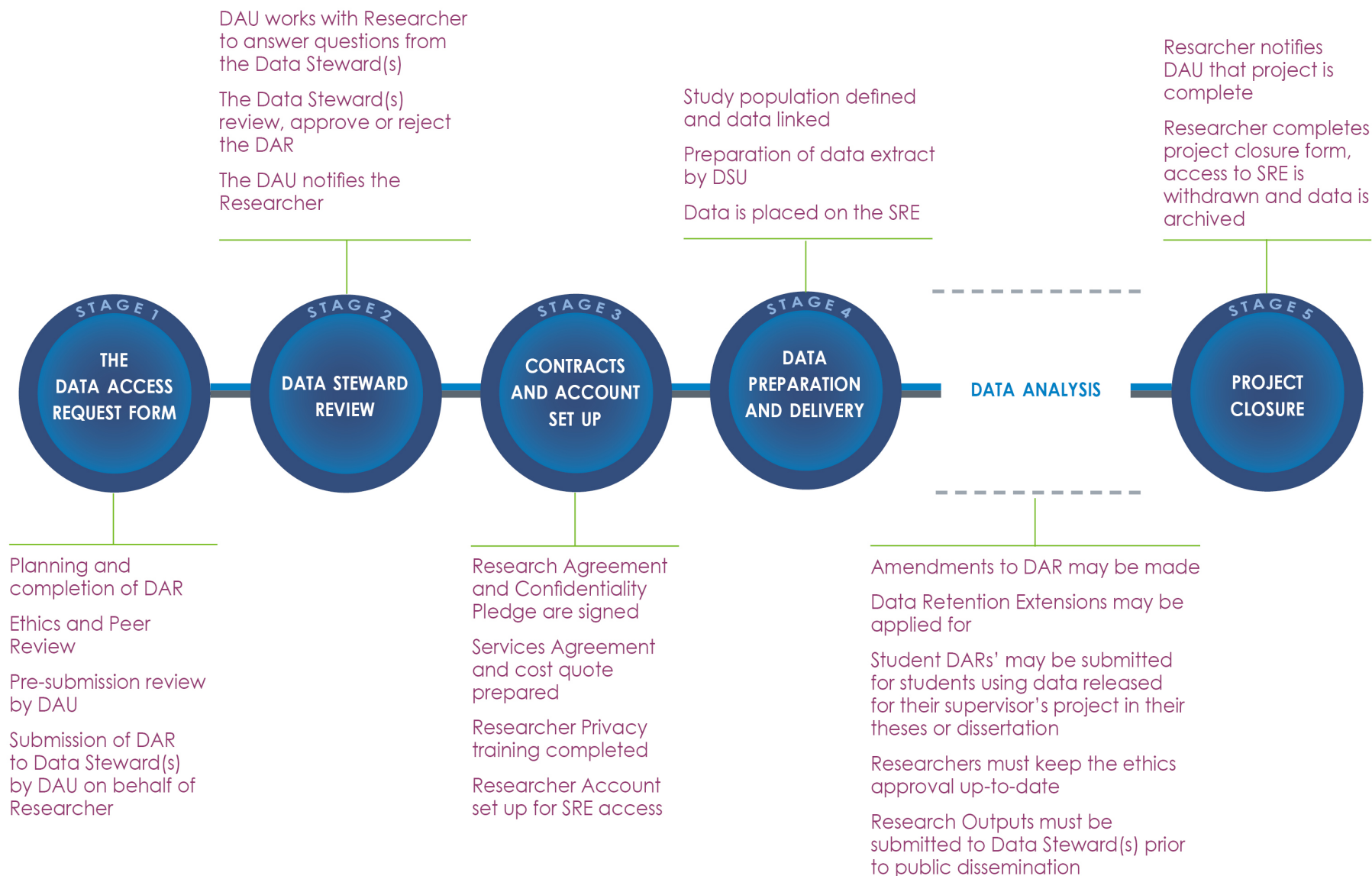


Linkage allows information on an individual from **one data source** to be **linked** to information on the **same** individual from **another data source**.



Linkage can also be used for **longitudinal analyses** within the **same data source**, even when identifying information is recorded inconsistently, incompletely or has changed over time.

THE DATA ACCESS PROCESS



THE DATA ACCESS PROCESS

Stage 4: Data preparation and delivery



WHAT HAPPENS?

Study population defined and data linked
Preparation of data extract by DSU
Data is placed on the SRE

DATA PROTECTION and PRIVACY

External audits and reviews

PRIVACY IMPACT ASSESSMENT

Developed, reviewed and approved by:

- UBC Office of the University Counsel
- BC Office of the Chief Information Officer
- BC Office of the Information and Privacy Commissioner

AUDITS

- Completed Systems and Security Audit in 2007 by BC Ministry of Health
- Expanded Systems and Security Audit in 2009 by Grant Thornton
- Detailed review of security and privacy controls and technical testing in 2014 by Deloitte

SITE VISITS

- David Flaherty, Former Privacy Commissioner for BC
- BC Office of the Chief Information Officer
- BC Office of the Information and Privacy Commissioner
- Data Stewards

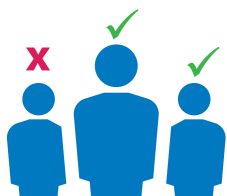
POLICIES & PROCEDURES

- Extensive privacy policies and procedures, regularly revised and updated
- Information Security Policies confirmed by Grant Thornton to be fully compliant with ISO/IEC 27002 requirements

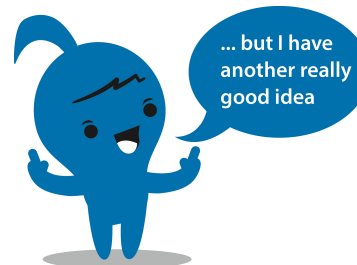
SECURE DATA STORAGE & ARCHIVING

Data protection and privacy

Procedural measures



Only named researchers have access to data



Data can only be used for requested purpose



Confidentiality pledges for research team



Pre-publication transcripts are reviewed by Data Steward(s)



Mandatory privacy training for researchers



Data is destroyed upon closure of project

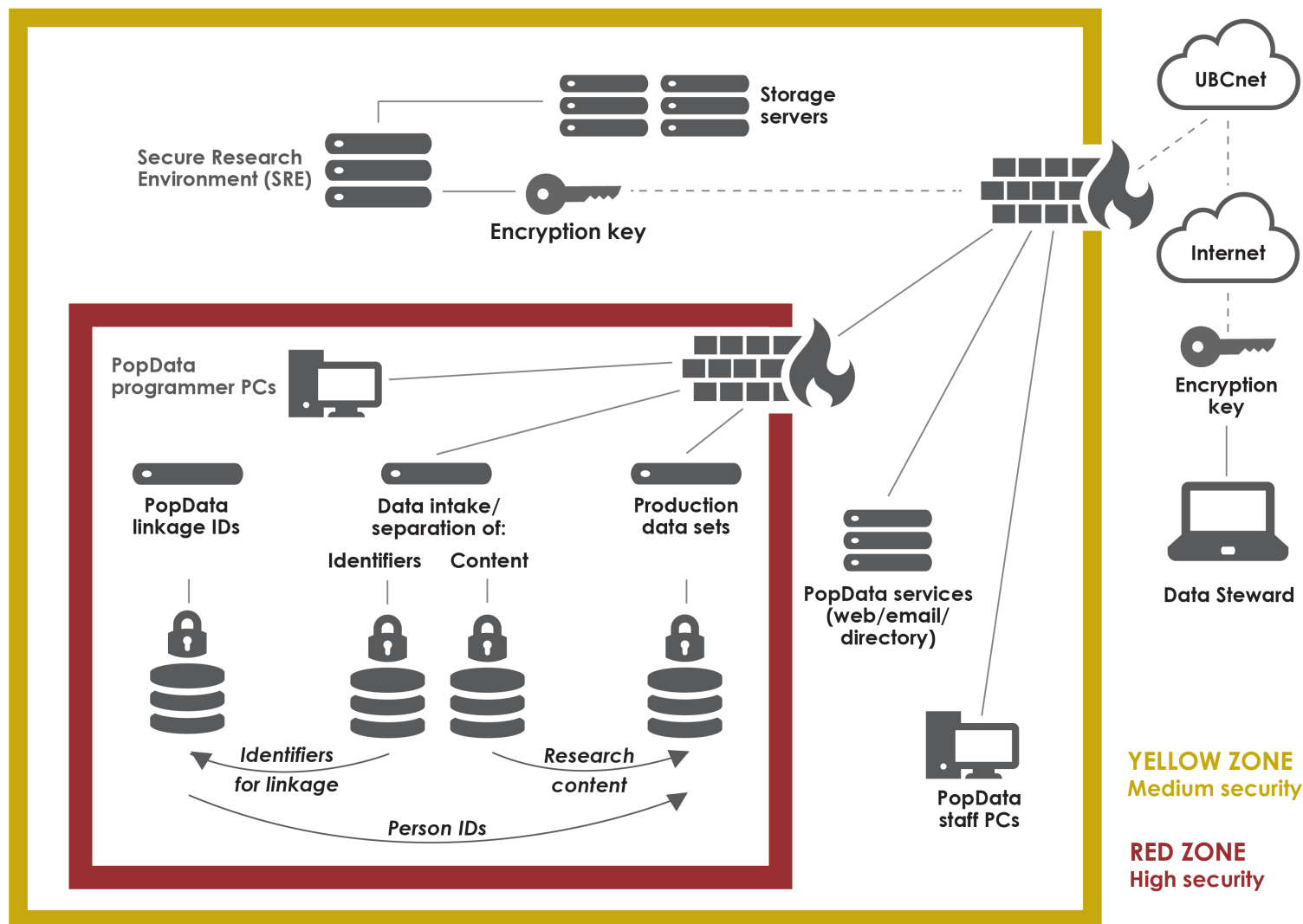
DATA PROTECTION and PRIVACY

Physical measures



DATA PROTECTION and PRIVACY

Network security

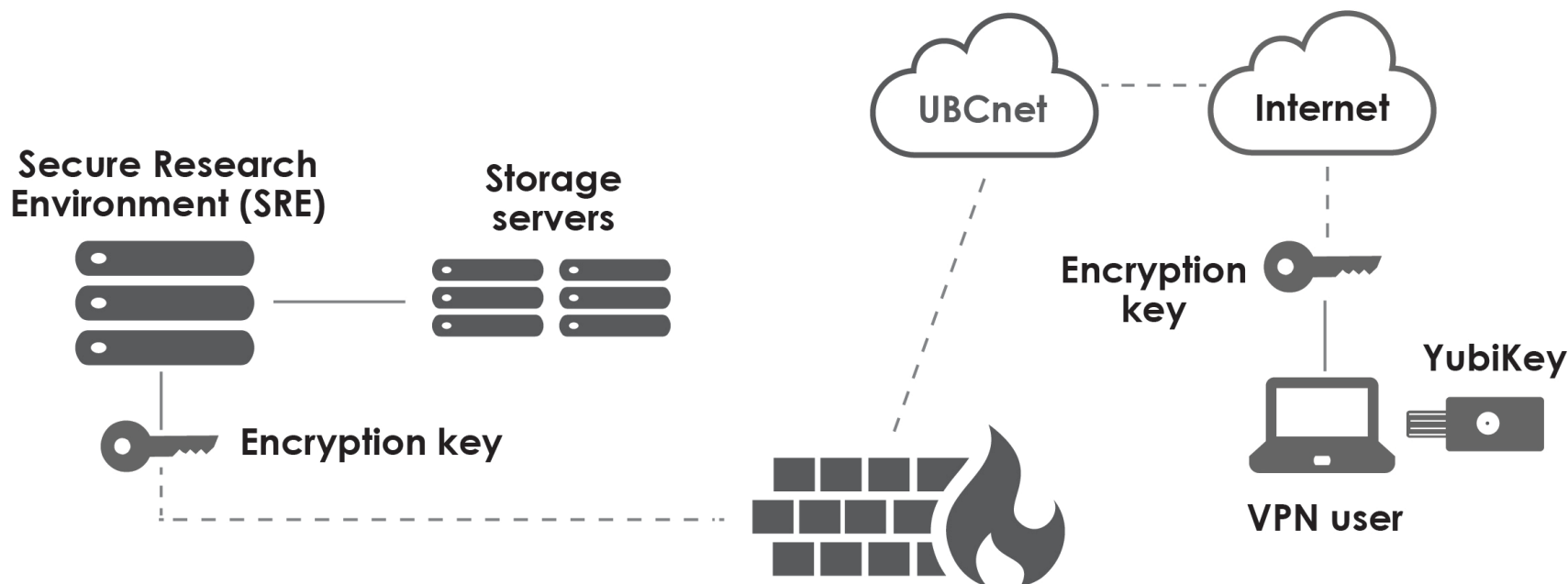


Receiving data on an encrypted CD





The Secure Research Environment



The SRE provides:

- Secure storage and backup
- Centralized location for access and processing of research data
- A range of software for use in data analysis
- Security standards that meet Data Steward(s) requirements

Trade Offs:



Researcher Perspective:

Data wanna
be free!!



The Costco Advantage:



PopData / Data Provider Perspective:



Security & Privacy Controls:

- Firewalled and restricted environment
- Two factor authentication
- Outputs limited and controlled
- Access can be turned off
- Data is encrypted on disk
- Secure backups
- Access can be restricted to Canada
- Audit controls
- Proper project closure

Sweetening the deal:



Taking Care of the Hassles:

- We take care of encryption
- We do the backups (offsite too)
- Hardware support
- Software upgrades
- Anti-virus
- Sharing within the team
- Project Archiving
- Many tools to choose from:



The Secure Research Environment

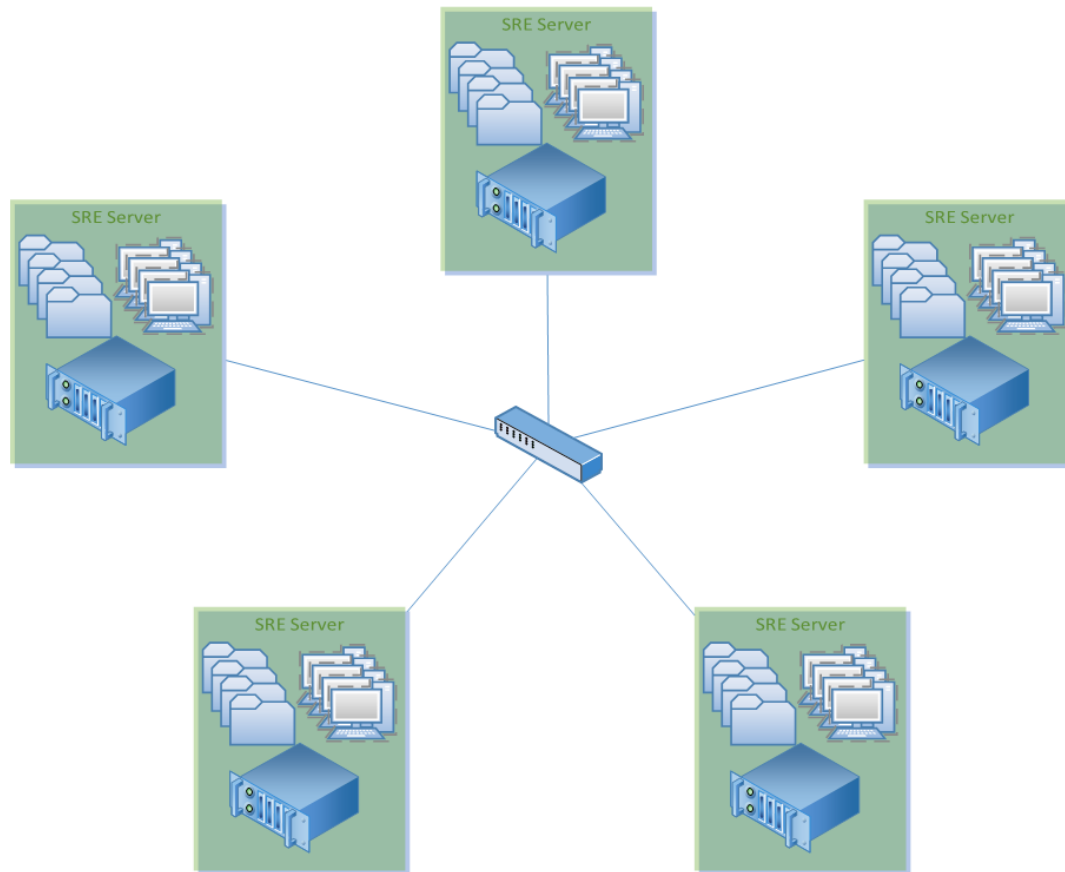
Software

- ACG John Hopkins 11.0.1
- ArcGIS Desktop 10.5.0.6491
- Epi Info 7.2.1
- GeoDa 1.8.14
- HLM 6.08
- JoinPoint 4.4.0.0
- MATLAB 2016b
- MLwiN 2.18
- MPlus 7.4
- MS Office 2013
- MS Project 2013
- MS Visio 2013
- Python 2.7.11
- Python 3.5.1
- R 3.3.2
- Rstudio 1.0.136
- SAS 9.4 TS Level 1M3
- SAS Entertainment Guide 7.13
- SaTScan 9.4.2 64-bit
- SPSS (IBM) Version 23
- SPSS Python Essentials Plugin
- SPSS Integration Plugin for R
- Stata 14.2
- WinBUGS 1.4.3

Utilities

- Adobe reader DC
15.023.20056
- Cygwin
- MS Visual FoxPro 9.0 SP2
- Notepad ++ 7.2.2
- Putty 6.07
- TextPad 7.6.2
- 7zip 16.04
- Java plugin Ver 8 Update 121
- Cute PDF Writer 3.1

SRE Architecture



Secret Ingredient:



SRE Architecture

- HP servers running Debian Linux
- KVM for virtualization
- Custom scripts for spinning up VMs from gold image
- Backing image with overlays
- 10Gbe network
- Login scripts mount the project R: drive

SRE Machines:

Regular Machines:

- Windows 7 (transitioning to Windows 10)
- Two virtual CPUs
- 16GB memory

Fast Machines:

- 32GB memory
- Quad-core Intel i7

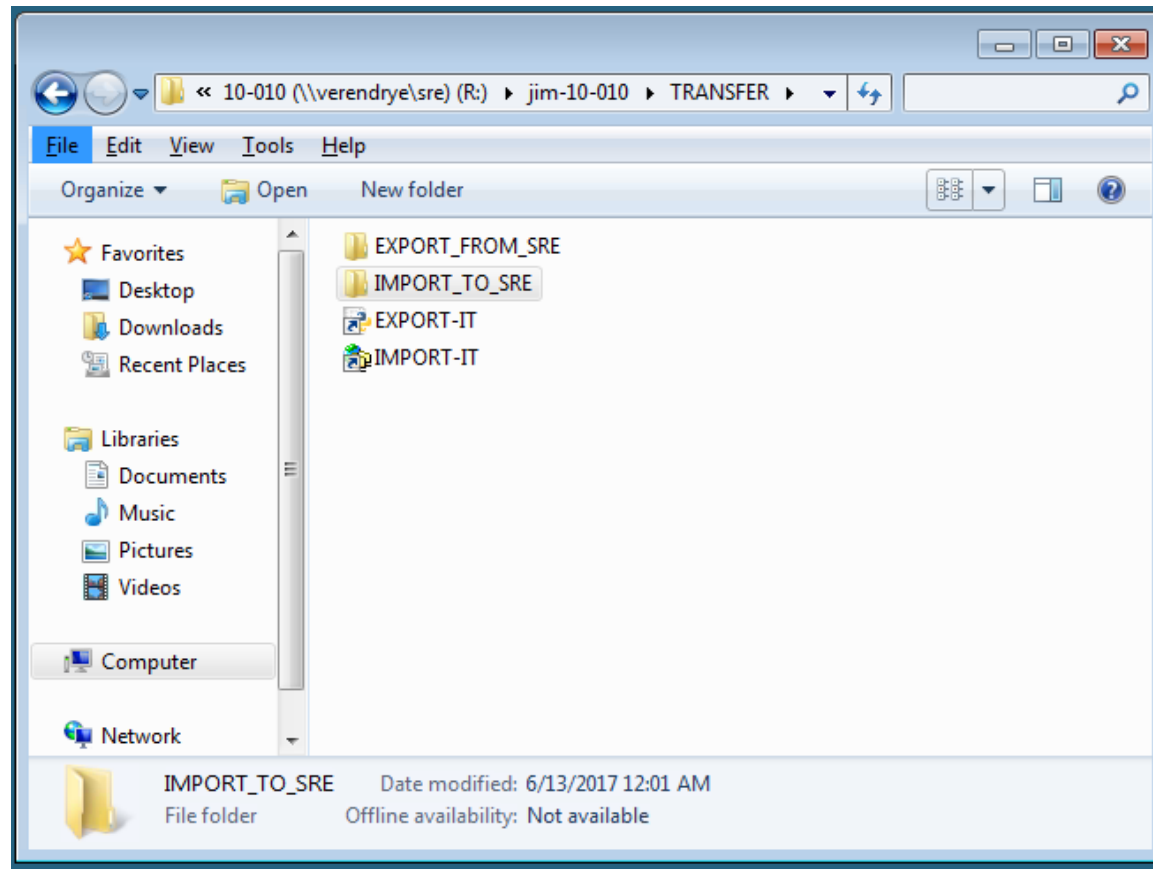
File Space:

- Each project has its own share
- Read only folder for the Data
- Separate private home directory for each project member
- Import and Export directories for each user
- Shared working directory
- 500GB initial quota, no hard quota, extra is charged

Importing and Exporting:

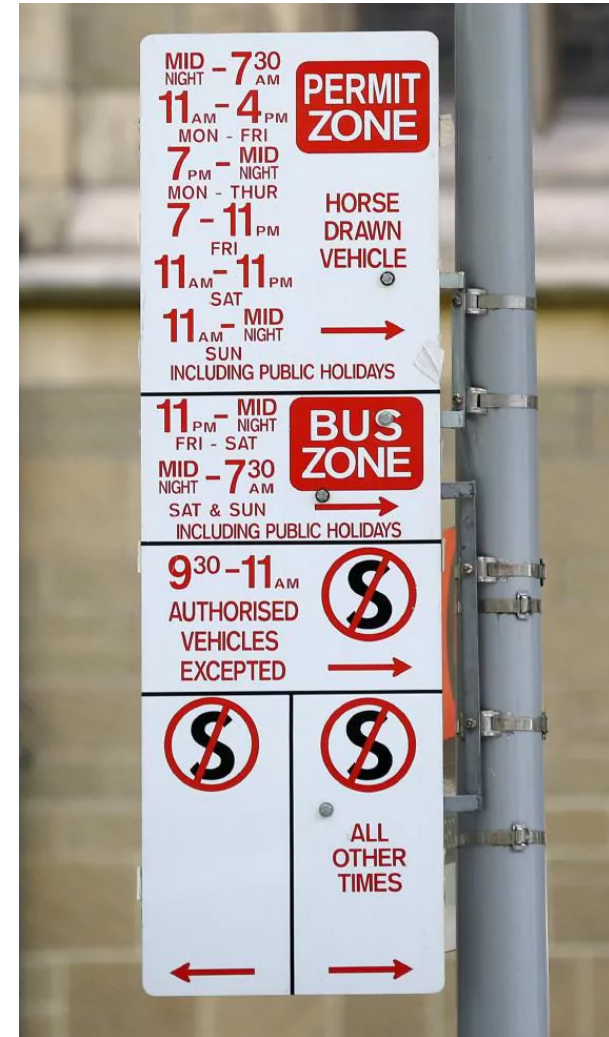


Yellowfolders:



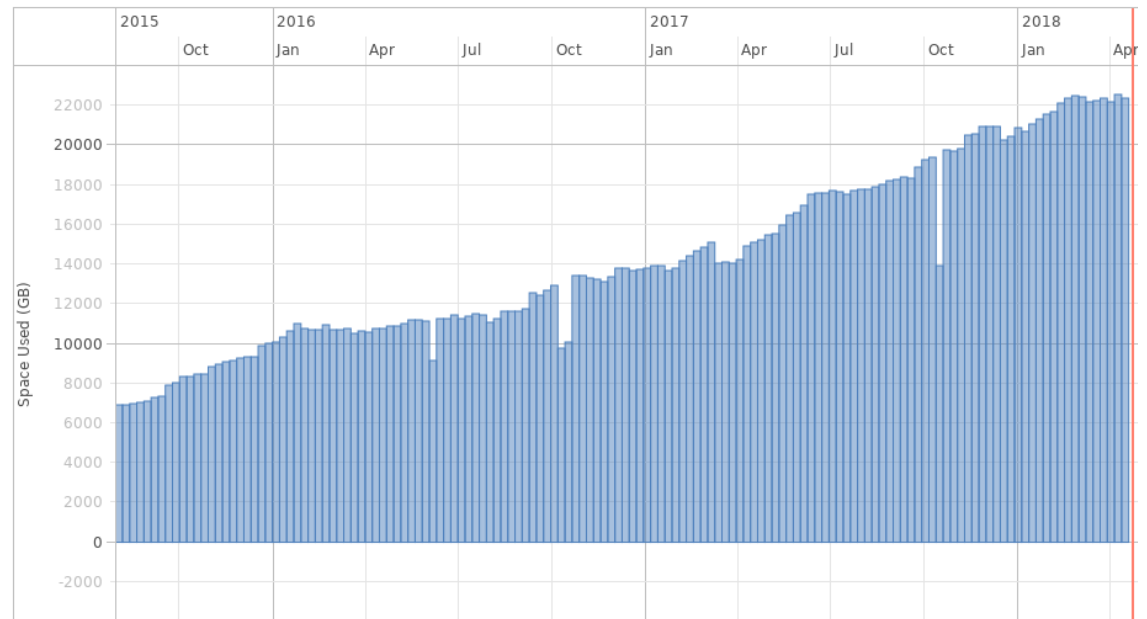
Restrictions:

- Size
- File type
- Small cell size (≤ 5)
- Archives
- Binaries
- Native stats program file types

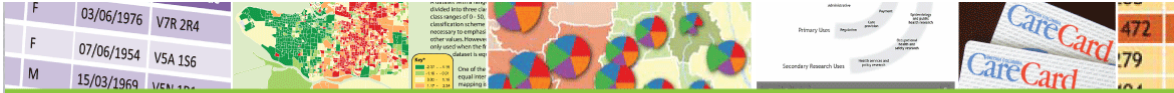



Problems / Issues:

- Camping out on Fast machines
- Data growth
- Specialized software



Snippets:





[Snippet List](#)
[Education & Training](#)
[Secure Research Environment](#)
[Researcher Tools](#)
[DSWG Info](#)
[Help & Quick Links](#)
[Account](#)

Welcome to PopData Snippets

PopData Snippets is a place where researchers and data analysts can share various bits or *snippets* of code they have written. Here you can find bits of code to help you read data into various statistical programs, sort data, and assist with various other commonly performed tasks.

The snippets are not meant as complete programs, or to cover every eventuality. They are just bits of code that you can add to your own programs to save you some work.

You can submit new snippets or use existing snippets both inside and outside the SRE. All snippets will be first reviewed and approved by Population Data BC staff. However, please keep in mind that Population Data BC will not be checking the snippets for usefulness, validity, or correctness. The snippets are provided **as is** and Population Data BC does not regularly monitor snippets or comments submitted by users post-approval. Don't use code that you haven't verified or don't understand.

Population Data BC makes no warranty, either expressed or implied, including, but not limited to, the warranties of merchantability and fitness for any particular purpose. The entire risk as to the quality of the code is with the user. In no event will Population Data BC be liable for any direct, special, indirect, consequential or other damages arising from users' use of codes on PopData Snippets. Population Data BC reserves the right to remove any code on the PopData Snippets without notice.

If you notice an error, please help us improve our library by providing comments on the snippet. Please note that comments will be moderated by Population Data BC and comments may be removed without notice.

[Add Snippet](#)

Tags

[Cleaning](#)
[Demographics](#)
[egen](#)
[flag](#)
[import](#)
[mean](#)
[merge](#)
[MSP](#)
[PNet](#)

[PNet DPD](#)
[Registry](#)
[sas](#)
[scan](#)
[spss](#)
[stata](#)

New Snippets

Title:	Language:	Created:
Create a rurality variable	Stata code	May 15, 2017, 3:44 p.m.
SAS Macro to scan through the data set and flag out the records as needed	SAS code	May 9, 2017, 12:15 p.m.
Excluding patients who are members of Federal Public Drug Benefit programs from PNet	SAS code	March 23, 2017, 5:44 p.m.
Importing Health Canada Drug Product Database (DPD) into SAS	SAS code	March 23, 2017, 4:47 p.m.
Importing chosen fiscal years of all zipped data into SAS using SAS macro (MSP, PNet, DAD,... etc)	SAS code	March 21, 2017, 1:05 p.m.

All Snippets

All Snippets Search:

Title:	Language:	Author:	Last Changed:	Tags:
Create a rurality variable	Stata code	Deirdre Weymann	2017/05/15 15:50	stata Demographics
Creating means by sub-category	Stata code	Melissa Medearis	2017/04/05 13:49	stata mean egen
Excluding patients who are members of Federal Public Drug Benefit programs from PNet	SAS code	Lucy Cheng	2017/03/29 09:17	PNet sas Cleaning
Importing chosen fiscal years of all zipped data into SAS using SAS macro (MSP, PNet, DAD,... etc)	SAS code	Lucy Cheng	2017/03/21 13:56	MSP import sas
Importing data into SPSS	SPSS code	Melissa Medearis	2017/03/22 09:10	import spss

Snippets:

[View Snippet](#)
[Education & Training](#)
[Secure Research Environment](#)
[Researcher Tools](#)
[DSWG Info](#)
[Help & Quick Links](#)
[Account](#)

Create a rurality variable

This Stata code classifies BC LHAs based on their degree of rurality.

The code creates and labels two categorical rurality variables. The first variable, "rurality_long", classifies LHAs into 9 categories: metro-core, metro-inner suburb, metro-outer suburb, mostly urban with large centre, mostly urban with medium centre, mixed urban/rural, rural <1hr from hospital, rural >1-<2.5 hrs from hospital, and rural >2.5 hrs from hospital. The second variable, "rurality", which is based on "rurality_long", classifies LHAs into 3 categories: urban/metro, mixed, and rural.

Snip Info

Author: Deirdre Weymann
Created: May 15, 2017, 3:44 p.m.
Changed: May 15, 2017, 3:50 p.m.
Language: Stata code
Tags: [stata](#) [Demographics](#)

[View Comments](#)
[Return to Snippets List](#)

Code:

```

1  * Rurality classification, long
2  gen rurality_long = 8
3  replace rurality_long = 0 if (lha==161|lha==162|lha==163|lha==164|lha==165|lha==166|lha==61)
4  replace rurality_long = 1 if (lha==40|lha==41|lha==38|lha==44|lha==45|lha==62|lha==63)
5  replace rurality_long = 2 if (lha==33|lha==34|lha==35|lha==37|lha==42|lha==43|lha==75|lha==201|lha==202)
6  replace rurality_long = 3 if (lha==23|lha==68|lha==24|lha==57)
7  replace rurality_long = 4 if (lha==22|lha==71|lha==15|lha==72)
8  replace rurality_long = 5 if (lha==65|lha==69|lha==70|lha==2|lha==60|lha==88|lha==48|lha==47|lha==28|lha==27|lha==52|lha==59|lha==64)
9  replace rurality_long = 6 if (lha==31|lha==77|lha==3|lha==14|lha==32|lha==21|lha==66|lha==76|lha==78|lha==16)
10 replace rurality_long = 7 if (lha==19|lha==5|lha==1|lha==12|lha==4|lha==17|lha==29|lha==56|lha==25|lha==30|lha==84|lha==6|lha==7)
11 replace rurality_long = . if lha==.
12 label define rurality_long 0 "Metro-core" 1 "Metro-inner suburb" 2 "Metro-outer suburb" 3 "Mostly Urban with large centre" 4 "Mostly Urban with medium centre" 5 "Mixed urban/rural" 6 "Rural <1hr from hospital" 7 "Rural >1-<2.5 hrs from hospital" 8 "Rural >2.5 hrs from hospital"
13 label values rurality_long rurality_long
14 * Rurality classification, condensed

```


AppTracker:

AppTracker

Jimbo (18-123)

Education & Training
Secure Research Environment
Researcher Tools
DSWG Info
Help & Quick Links
Account

Main
Project Info
Project Activity
Data Preparations

Main:

Title:

Coping Strategies: The Timing and Impact of Pizza Deliveries

Current Status:

APPROVED-ALL-DATA-RECEIVED on 2018-03-05

Student DAR:

No

List of Data Steward Organizations:

- College of Physicians and Surgeons of BC (CPSBC)
- Ministry of Health (MoH)

Parties:

Applicants

Coordinators

Co-Investigators

Additional Team Members

Data Retention:

2019-03-31	College of Physicians and Surgeons of BC (CPSBC)
2019-01-31	Ministry of Health (MoH)

Data Storage Location

Approved Data Storage:

SRE

Invoice No.:

20J45215-79J

SRE Expiry Date:

2019-01-30

Ethics

Certificate No.	Expiry Date
J09-12977	2019-02-20



AppTracker:

AppTracker

Jimbo (18-123)

Education & Training ▾

Secure Research Environment ▾

Researcher Tools ▾

DSWG Info

Help & Quick Links ▾

Account ▾

Main

Project Info

Project Activity

Data Preparations

Project Activity Summary

Number of Amendments: 3

Number of Data Retention Extensions: 5

Number of Student DARs: 0

2018-01-09: data-retention-extension ⚙

2016-12-07: data-retention-extension ⚙

2016-01-22: data-retention-extension ⚙

2016-01-19: add-member ⚙

2015-02-04: data-retention-extension ⚙

2014-10-06: add-member ⚙

Reference No: M-18-123-898

Type: ADD-MEMBER

Description: Add Bob McKenzie (data access) and Doug McKenzie (data access)

Full Approved Date: Oct. 6, 2014

add-member details:

Data Steward Organization	Submit Date	Approved Date	Notes
Ministry of Health (MoH)	Oct. 1, 2014	Oct. 2, 2014	
College of Physicians and Surgeons of BC (CPSBC)	Oct. 3, 2014	Oct. 6, 2014	

2014-01-21: amendment ⚙

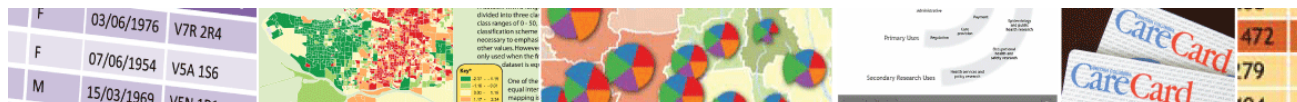
2013-11-05: amendment ⚙

2013-11-04: amendment ⚙

2013-09-17: data-retention-extension ⚙

2012-03-23: application ⚙

MetaData Central:



DARonline

 my.popdata

[Education & Training](#)
[Secure Research Environment](#)
[Researcher Tools](#)
[DSWG Info](#)
[Help & Quick Links](#)
[Account](#)

METADATA CENTRAL CONTENTS

Welcome to Population Data BC's MetaData Central (MDC)! The goal of MDC is to provide the research community that Population Data BC serves (i.e., researchers, data stewards, and PopData staff) with sufficient information to facilitate informed decisions about Population Data BC's holdings. Please remember that the information that you view here is not to be reported or reproduced for any reason. You will find MDC can function as a resource for file-level summary information (e.g., inclusions, exclusions, and fields available) as well as field-level information (e.g., field descriptions, background information, and basic descriptive statistics). This information can be found by selecting one of the files currently listed on MDC below.

HEALTH CARE AND HEALTH SERVICES DATA

[Discharge Abstracts Database \(Hospital Separations\) \(April 1, 1985 onwards\)](#)

[Medical Services Plan Payment Information File \(April 1, 1985 onwards\)](#)

[Home and Community Care \(January 1, 1990 onwards\)](#)

[BC Perinatal Data Registry \(April 1, 2000 to March 31, 2014\)](#)

POPULATION AND DEMOGRAPHIC DATA

[Consolidation File \(January 1, 1986 onwards\)](#)

[Vital Statistics Births \(January 1, 1986 onwards\)](#)

[Vital Statistics Deaths \(January 1, 1985 onwards\)](#)

[Vital Statistics Marriages \(January 1, 1986 onwards\)](#)

[Vital Statistics Stillbirths \(January 1, 1986 onwards\)](#)

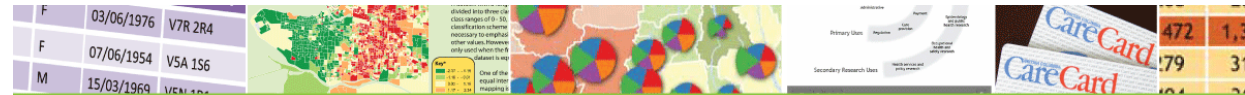
EARLY CHILDHOOD DATA

[Early Development Instrument \(EDI\)](#)

OTHER

[PharmaCare \(January 1, 1985 onwards\)](#)

MetaData Central:



DARonline

my.popdata

Education & Training

Secure Research Environment

Researcher Tools

DSWG Info

Help & Quick Links

Account

[Metadata Central Contents](#) -> [Vital Statistics Births \(January 1, 2036 onwards\)](#)

Sex

Sex of Child.

Detailed Description

Sex of Child.

Background Information

None documented.

Basic Quality Issues

None documented.

Code Values

M = Male;
F = Female;
O = Other;
U = Martian

+ Show Help

Warning The information and values provided are, under no circumstances, to be reported or reproduced in any manner. The following descriptive statistics are for reference only.

Below is a crude field summary which does not account for other field or record contexts. The data represented is not static and subject to revisions. The information provided is not suitable for comparison with released data, official reports or publications.

FREQUENCY_SUMMARY

Value	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	20
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	210,848	280,599	20,440	2,121	21,405	22,366	222,267	22,619	62,449	22,790	22,681	29,379	21,797	21,079	270,313	19,802	19,731	19,393	19,744	19,713	19
M	212,394	231,553	261,507	272,086	22,515	23,368	233,457	234,690	23,767	24,338	24,276	23,823	122,819	21,179	21,660	20,890	220,822	20,723	230,766	20,816	20
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U	0	15	15	0	15	0	15	0	15	15	15	0	15	15	0	0	0	0	15	15	0
Total	43,242	412,154	41,948	63,207	43,922	45,734	15,726	46,309	46,216	47,130	96,958	46,202	44,618	9,059	41,973	40,692	40,553	40,116	940,512	10,530	40

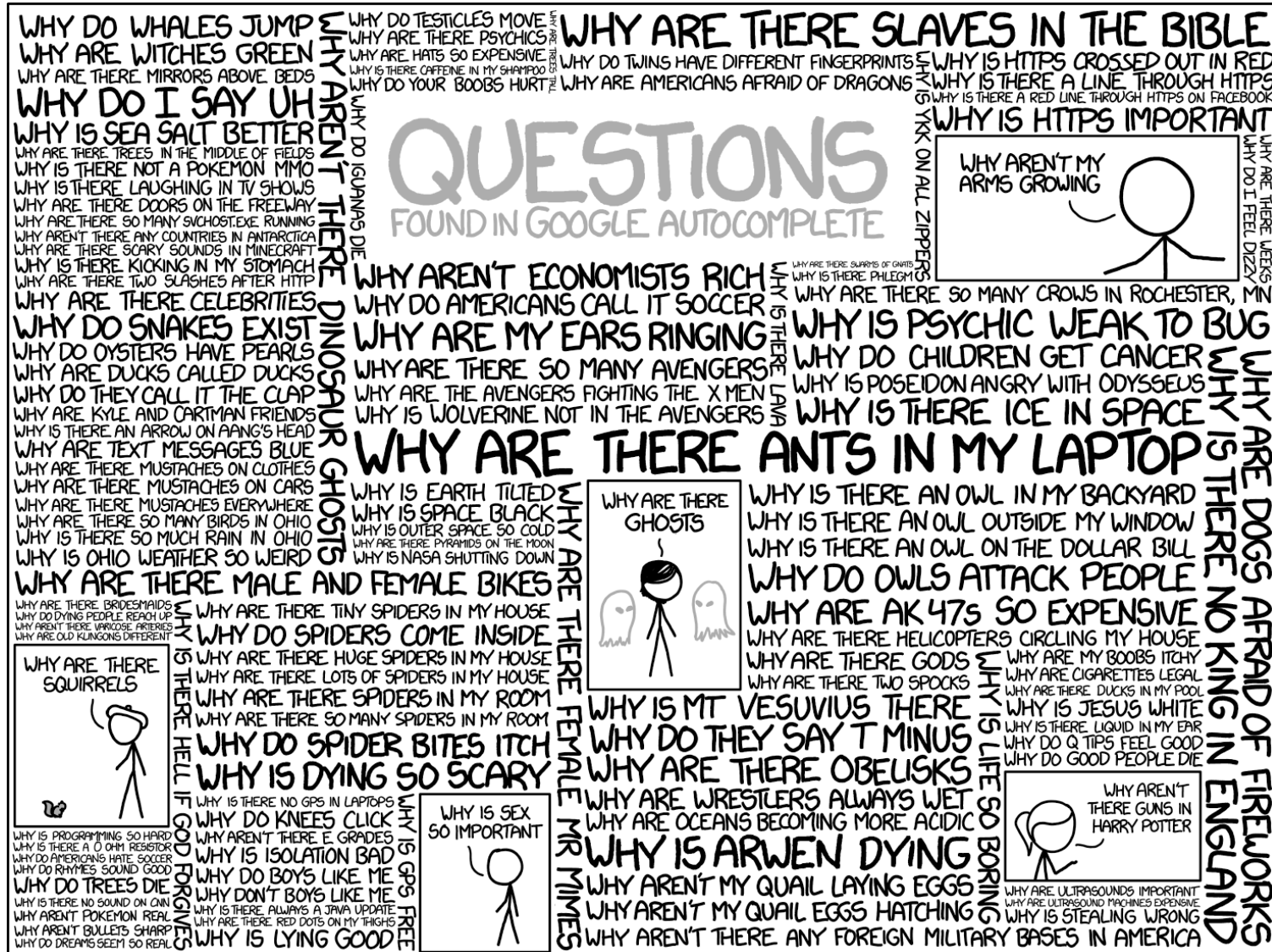
Available Summaries:

[c_sex - FREQUENCY_SUMMARY](#)
[c_sex - PERCENTAGE_SUMMARY](#)

Available Frequency Databases:

☒ births2048-08-14_freq
☐ births2045-10-26_freq
☐ births2041-11-27_freq
☐ births2037-10-10_freq

Questions?





populationdata^{BC}

A PLATFORM SUPPORTING RESEARCH ON
HUMAN HEALTH, WELL-BEING AND DEVELOPMENT.

CONTACT INFORMATION

Jim Mintha

Population Data BC

604-827-3378

jim@popdata.bc.ca

www.popdata.bc.ca

