populationdata BC

**Linking data for research:
Informing policy-making for healthier communities**

September 2024

Population Data BC (PopData)
Our goals

1. Innovative practices

To create dynamic, adaptable and secure data access management processes that meet the expectations of the public, data providers, and data users

2. Problem-solving

To provide solutions that respond to the challenges experienced by BC's public, private and non-profit sectors in the collection, management and/or analysis of complex data
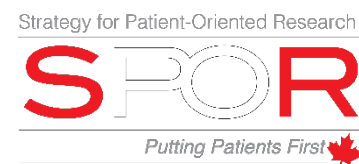
3. Partnerships

To collaborate to mobilize data and leverage the expertise of PopData and our partners, to shape the development of innovative programs and policies for the public good

PopData
Our funders (former, present and in kind)

Former and foundational

BC Ministry of Small Business, Technology and
Economic Development Knowledge Development Fund

PopData services for Researchers

Data access

- Guidance on which of 3 data access streams is best for your project
- Support for preparation and submission of your data access request
- Coordination and management of PopData/HDPBC projects
- Range of online tools and resources

Data linkage

- PopData projects
- Data Innovation Program projects

Secure data storage

- PopData projects
- Data Innovation Program projects
- CaraSpace projects

Education and training

- Professional Specialization Certificate in Population Health Data Analysis

Milestones in the evolution of PopData

populationdata BC

1996
Researchers began using the British Columbia Linked Health Database (BCLHD)

2005
Population Health Learning Observatory (PHLO) established

2008
Population Health Learning Observatory (PHLO) became **Population Data BC**

2009
Secure Research Environment launched

2011
Launch of PHDA certificate program in collaboration with the University of Victoria

2012
Became Director of International Health Data Linkage Network (now International Population Data Linkage Network (IPDLN)) for 2013-2014 term

2016
Became part of the BC SUPPORT Unit

2018
Partnered with BC government's Data Innovation Program providing data access for government projects

2019
Founder, member and network administration host of Health Data Research Network (HDRN) Canada supporting multi-regional research

2021
Expansion of Data Innovation Program to include academic projects

2024
Partnered with BC Ministry of Health to support data access through Health Data Platform BC (HDPBC)

What data are available for academic research?

Population coverage of data available to academic researchers

British Columbia has **5.6 MILLION** residents

Data is available on **ALL** of them **100% coverage**

Data available relate to, for example:

- H Hospital stays
- Student assessments
- Visits to a Doctor
- Medications prescribed
- Births, marriages & deaths
- Citizenship
- Pregnancy & childbirth
- Disease registries
- Workplace injuries
- ICBC - Driver Licensing

For some people there are **a lot of data** for other people **very little**

How much data is available on a person depends on their **life experience**

Population coverage of data available to academic researchers
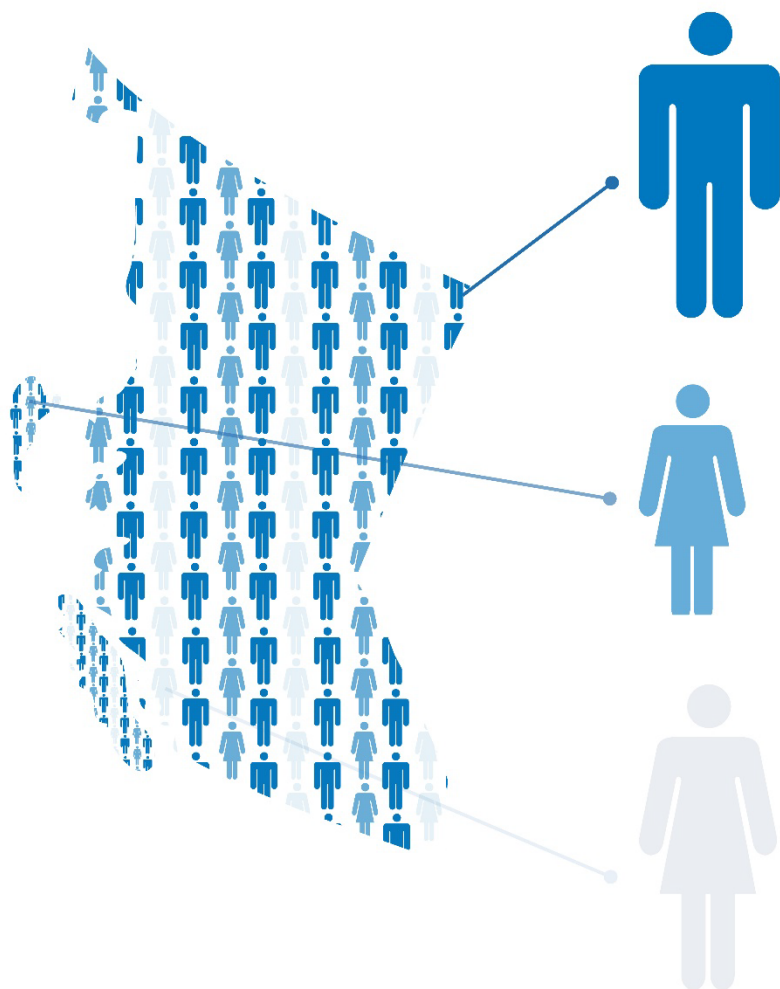
26 YEAR OLD MALE

Born in France
Canadian Citizen at 9
Has high blood pressure
Broke leg at work
Married

6 YEAR OLD FEMALE

Born in BC
Attends elementary school

66 YEAR OLD FEMALE

Retired to BC in 2015
A few physician visits

Accessing BC data for academic research

There are currently **three streams** available for academic researchers to access data on the population of British Columbia:

- **Population Data BC (PopData)**
- **The Data Innovation Program (DI Program)**
- **Health Data Platform BC (HDPBC)**

Researchers will be routed to a data access stream based on a range of qualifying criteria which includes:

- Student or non-student status
- The data required to answer the research question
- The tools required for analysis

Data sets available through PopData stream

populationdata BC

HEALTH
14 data sets

BC Cancer Registry

BC Perinatal Data Registry

Canadian Longitudinal Study on Aging

Chronic Disease Registry

COVID-19 Immunization Data

COVID-19 Test Lab Data

Discharge Abstracts Database

Home and Community Care

Medical Services Plan

Mental Health Services

National Ambulatory Care Reporting System

Patient-Centered Measurement Surveys

PharmaCare

PharmaNet

Residential Assessment Instrument

WORK & INCOME
2 data sets

Income Band

Claims and Firm Level Files

EDUCATION
3 data sets

K to 12 Student Assessments

K to 12 Student Demographics and Achievements

K to 12 Student Learning Survey

ENVIRONMENT & RESOURCES
multiple data sets

Canadian Urban Environmental Health Research Consortium

SOCIAL
3 data sets

Early Development Instrument

Middle Years Development Instrument

Permanent Residents

DEMOGRAPHIC
6 data sets

Central Demographics File

Registration and Premium Billings

Vital Events and Statistics: Births, Stillbirths, Marriages and Deaths

Data sets available through the Data Innovation Program stream

HEALTH - 11 data sets

BC Perinatal Data Registry
Chronic Disease Registry
COVID-19 Test Lab Data
Discharge Abstracts Database
Home and Community Care
Medical Services Plan
Mental Health Services
National Ambulatory Care Reporting System
PharmaCare
PharmaNet
Residential Assessment Instrument

EDUCATION - 4 data sets

K to 12 Student Assessments
K to 12 Student Demographics and Achievements
K to 12 Student Learning Surveys
K to 12 Socio Economic Index

SOCIAL - 11 data sets

Affordable Child Care Benefit
BC Employment and Assistance
BC Family Maintenance
Child and Youth Mental Health
Child Care Subsidy
Child Welfare Program
Community Living Programs
Early Development Instrument
Private Market Rent Supplements for Eligible Seniors and Working Families
Shelter and Homeless Outreach Private Market Rent Supplements
Supportive Housing

TRANSPORTATION - 5 data sets

ICBC – Contraventions
ICBC – Crashes
ICBC – Driver Licensing
ICBC – Traffic Accident System
ICBC – Vehicle Population

DEMOGRAPHIC - 7 data sets

BC Population Projections
Central Demographics File
Registration and Premium Billings
Vital Events and Statistics: Births, Stillbirths, Marriages and Deaths

JUSTICE - 1 data set

Corrections – Adult Community and Custody

WORK & INCOME - 5 data sets

Career Paths for Skilled Immigrants
Income Bands – Custom
Income Bands by Postal Code
Labour Market Programs
Provincial Nominee Program

Data sets available through the HDPBC stream

![populationdata BC logo]

HEALTH
23 data sets

BC Cancer Registry

BC Cardiac Services Registry

BC Perinatal Data Registry

BC Renal

Chronic Disease Registry

CIHI Population Grouping Methodology and CPOP Diagnosis Codes

COVID-19 Hospitalization and Critical Care Reports

COVID-19 Test Lab Data

Discharge Abstracts Database

HealthIdeas Common and Reference Dimensions

HealthLink 811-COVID

Health System Matrix

Home and Community Care Minimum Reporting Requirements

Medical Imaging Wait Times

Medical Services Plan

National Ambulatory Care Reporting System

Patient Centered Measurement Surveys

PharmaNet

Post-Covid Recovery Clinic

Provincial Immunization Registry

Residential Assessment Instrument

Surgical Wait Times

VPP Clinical OMOP Data

DEMOGRAPHIC
3 data sets

Client Roster

Population Extrapolation for Organizational Planning with Less Error (PEOPLE)

Vital Events: Births, Stillbirths and Deaths

SOCIAL
1 data set

Canadian Index of Multiple Deprivation

Comparing the 3 streams

	DI Program	PopData	HDPBC
Can data sets be linked to external data sets (including Researcher-collected data?	✓	✓	✓
Are students allowed to access the data?	☒	✓	✓
Are sensitive variables available?	✓	✓	✓
Can I choose/customize the variables in a data set? **Core Data**: Some BC Ministry of Health data sets available through PopData and HDPBC are called 'Core Data'. A Core Data set is a standardized 'bundle' of commonly requested variables. **Non-Core Data**: Non-Core Data are variables that are **NOT** included in the standardized Core Data set. Non-Core Data is available for request **as an addition** to the Core Data set.	Variables for the entire data set are provided (sensitive variables are masked)	Core (standard variables) Non-Core (sensitive variables) available by request	Core (standard variables) Non-Core (sensitive variables) available by request

PLEASE NOTE: The tools available for statistical analysis differ between the 3 streams.
Go to: www.popdata.bc.ca/data_access to check which stream has the tools you need.

Need data for a Pan-Canadian project?
HDRN Canada's Data Access Support Hub can help

Data Access Support Hub (DASH)

A one-stop data access service portal for researchers requiring multi-regional health and health-related administrative data in Canada.
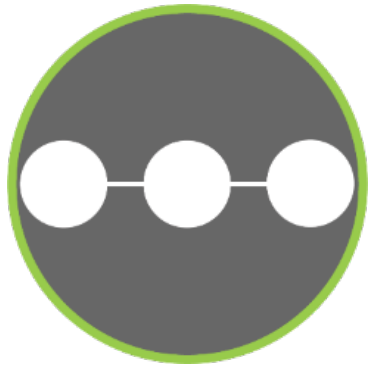
DASH streamlines the data access process through its centralized data access request (DAR) form and harmonizes processes to request data from multiple data repositories across the country.

Go to: https://www.hdrn.ca/en/dash/about-dash/

PopData is founder, member and administration host

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Data access

The data access process for PopData/HDPBC projects

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7
Planning the DAR	Completing & submitting the DAR	Data Steward Review	Contracts, accounts & privacy training	Data provisioning & release	Data analysis & Pre-publication review	Project closure

For information on the data access process for DI Program projects, go to:
https://www2.gov.bc.ca/gov/content/data/finding-and-sharing/data-innovation-program

The data access process for PopData/HDPBC projects
Stage 1: Planning the Data Access Request (DAR)

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7
Planning the DAR	Completing & submitting the DAR	Data Steward Review	Contracts, accounts & privacy training	Data provisioning & release	Data analysis & Pre-publication review	Project closure

What happens during this stage?

- Identify the data that you require
- If you require external data, contact the provider to discuss the feasibility of your request
- Request a feasibility and cost estimate
- Review BC Ministry of Health consent form requirements and draft consent forms (if applicable)
- Plan DAR using:
  - Support from PopData's Data Access Unit (DAU)
  - Online tools and resources

The data access process for PopData/HDPBC projects
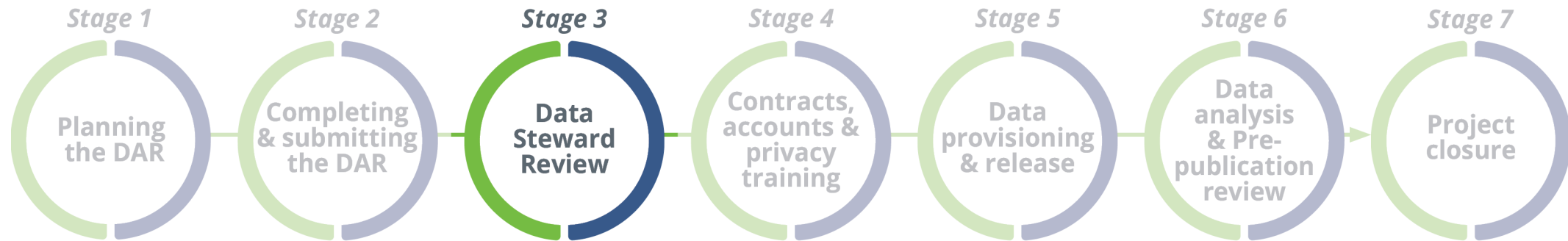## Stage 2: Completing and submitting the DAR

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7
Planning the DAR	Completing & submitting the DAR	Data Steward Review	Contracts, accounts & privacy training	Data provisioning & release	Data analysis & Pre-publication review	Project closure

What happens during this stage?

- Completion of Ethics and Peer Review
- Completion of consent form requirements to attach to DAR (if applicable)
- Completion of DAR via DARonline
- Co-ordination of the DAR and pre-submission review by Data Access Unit (DAU)
- Submission of DAR to Data Steward(s) by DAU on behalf of the Researcher

The data access process for PopData/HDPBC projects
Stage 3: Data Steward Review

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7
Planning the DAR	Completing & submitting the DAR	Data Steward Review	Contracts, accounts & privacy training	Data provisioning & release	Data analysis & Pre-publication review	Project closure

What happens during this stage?

- DAU works with Researcher to answer any questions from the Data Steward(s)
- The Data Steward(s) review, approve or reject the DAR
- The DAU notifies the Researcher

The data access process for PopData/HDPBC projects
Stage 4: Contracts, accounts and privacy training

What happens during this stage?

- Research Agreement and Confidentiality Pledge are signed
- Services Agreement and Cost Quote are prepared
- Researcher accounts are set up to access Secure Research Environment (SRE) or HPDBC Secure Environment
- *Privacy Training 101 for Researchers* course is completed by the PI and all project team members requiring access to the data

The data access process for PopData/HDPBC projects
Stage 5: Data provisioning and release

Stage 1	Stage 2	Stage 3	Stage 4	**Stage 5**	Stage 6	Stage 7
Planning the DAR	Completing & submitting the DAR	Data Steward Review	Contracts, accounts & privacy training	**Data provisioning & release**	Data analysis & Pre-publication review	Project closure

What happens during this stage?

PopData projects

- Study population defined and data linked
- Preparation of data extract
- Data is placed in the Secure Research Environment (SRE) for analysis

HDPBC projects

- Data is prepared and provisioned
- Researchers create their own cohort or request support
- Authorized research team members receive access to the approved data via HDPBC Secure Environment.

The data access process for PopData/HDPBC projects
Stage 6: Researcher analysis and Pre-publication Review

Stage 1 — Planning the DAR

Stage 2 — Completing & submitting the DAR

Stage 3 — Data Steward Review

Stage 4 — Contracts, accounts & privacy training

Stage 5 — Data provisioning & release

Stage 6 — Data analysis & Pre-publication review

Stage 7 — Project closure
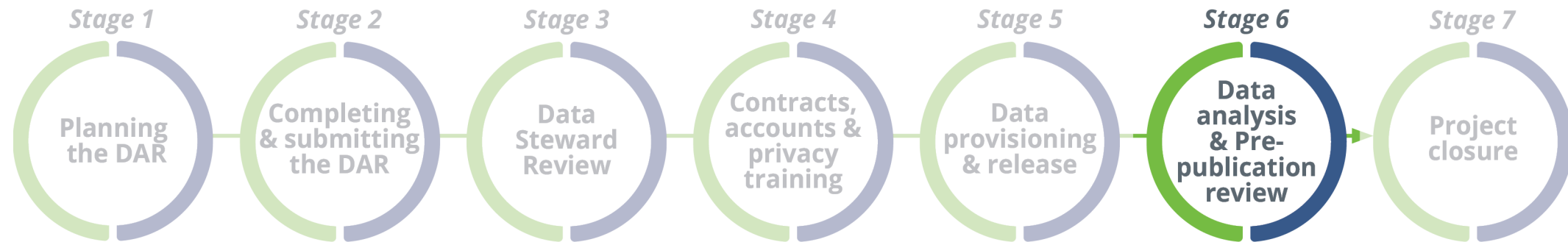
What happens during this stage?

- Researchers analyse data in the PopData Secure Research Environment or the HDPBC Secure Environment
- New study team members may be added
- Amendments to DAR may be made
- Data Retention Extensions may be applied for

- Researchers must keep ethics approval up-to-date
- Student DARs may be submitted to use data released under their supervisor's project for their theses or dissertation
- Research Materials must be submitted to Data Steward(s) prior to public dissemination

The data access process for PopData/HDPBC projects
Stage 7: Project closure

What happens during this stage?

- Researchers notify DAU that the project is complete and access to the data is no longer required
- Researcher completes project closure documentation
- Access to PopData Secure Research Environment or HDPBC Secure Environment is withdrawn
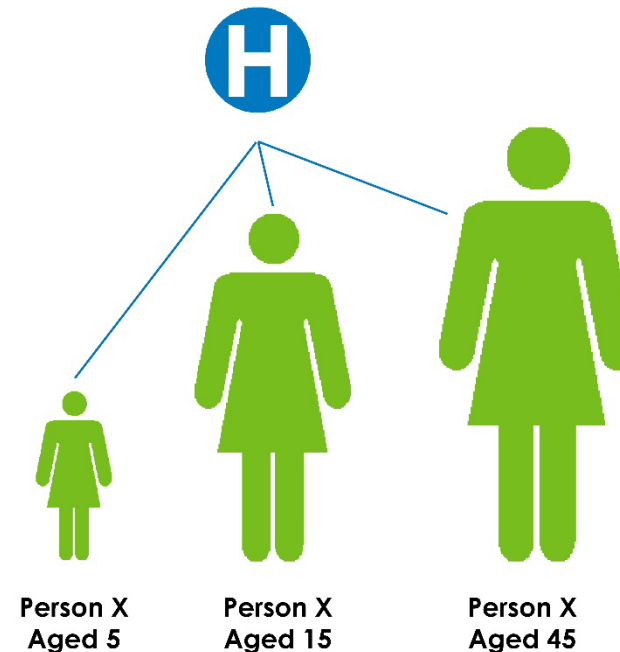- Data is archived for 7 years

Data linkage

Data linkage

Person X

- Perinatal record
- Workplace injury record
- Income band record
- Marriage record
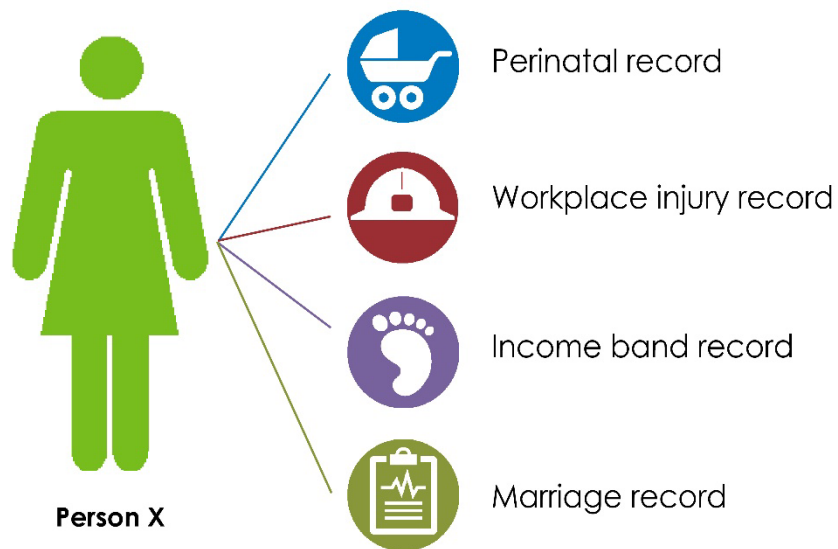
Linkage allows information on an individual from **one data source** to be **linked** to information on the **same** individual from **another data source**.

Hospital records

Person X Aged 5   **Person X Aged 15**   **Person X Aged 45**

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.

Data linkage
Facilitating linkages in a secure and privacy-sensitive manner

It all starts with … The Population Directory
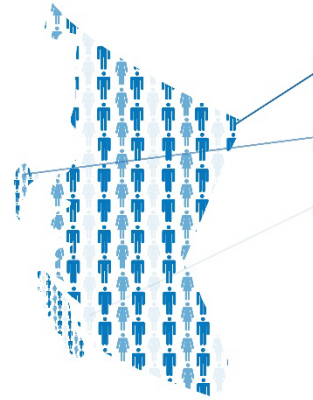
Name	DoB	Address	PHN	Linkage ID
A Person	12/06/1987	46 One Street, Town, City, Postal Code	123456	873595
B Person	03/07/1962	1 Two Crescent, Town, City, Postal Code	234567	456746
C Person	30/02/1993	6 Six Road, Town, City, Postal Code	7639574	238504

Then …

Stage 1
Separation of identifying information from content

Stage 2
Data cleaning

Stage 3
Ensuring accuracy

Stage 4
Resolution

Secure data storage

Secure data storage
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

Secure data storage
Technical measures for data protection and privacy

Fully compliant with recognized ISO requirements for information security. Network controls in place include:

- Firewall protection

- Access to Red Zone network, which holds the data, requires dual-factor authentication

- Access to network is restricted to named personnel

- All access is logged and audited

- Content data and identifier data are stored separated from each other in encrypted logical areas

- Red Zone networks are logically moated and have no direct connection to outside networks

- Encryption of all stored data including Personal Information (physical removal of the disks renders them unreadable)

Secure data storage
The Secure Research Environment (SRE)
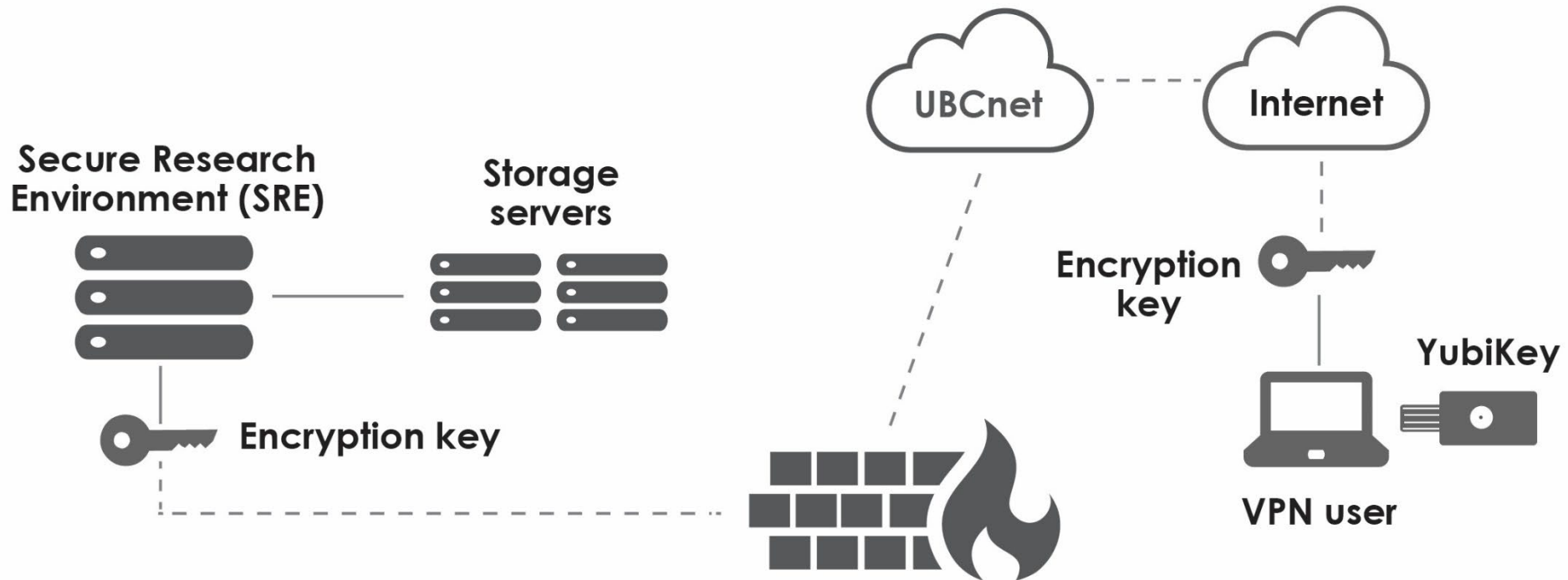
A secure private cloud accessible only via an encrypted Virtual Private Network (VPN) through a firewall and two-factor authentication.

Provides a safe, secure place for researchers to store and analyse their data, while ensuring that privacy-sensitive data remains protected.

Secure data storage
The Secure Research Environment (SRE)

Researcher features:

- Access to over 100 well-provisioned virtual machines

- A large suite of free analytical software

- Worry free management of data backup, antivirus protection, and other related computer maintenance

- Remote access from any computer or operating system

- Ability to share data and code between project members

Security features:

- Data are stored on encrypted fast solid state disks

- Projects are partitioned, users see only their own project data

- Continuous monitoring of all systems, network, and imports/exports

- Accounts/projects can be disabled or suspended at project closure, ethics expiry, or in the event of a privacy incident

Secure data storage
Procedural measures for data protection and privacy
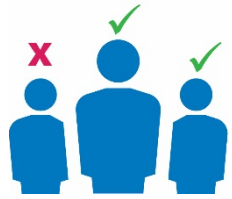
Only named researchers have access to data

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 archived then destroyed upon closure of project

Education and training

Education and training
Professional Specialization Certificate in Population Health Data Analysis
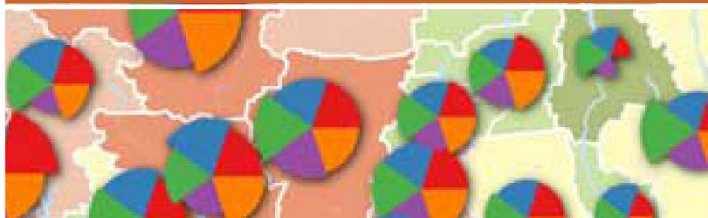
populationdata BC

**Linking data for research:
Informing policy-making for healthier communities**

www.popdata.bc.ca