# Data Innovation Program
## PRIVACY AND SECURITY FRAMEWORK

BRITISH COLUMBIA | Ministry of Citizens' Services

## About the Data Innovation Program

The Data Innovation (DI) Program is a data integration and analytics program for government. While every B.C. ministry collects and manages its own data, the program:

- Collects and links data from multiple ministries,
- De-identifies it, and
- Makes this de-identified data available to government analysts in a secure platform to support population-level research projects.

Although the B.C. government has always integrated data, the DI Program provides a more timely, secure and consistent way to do this work – making it easier for government to identify trends, respond to emerging issues, and develop solutions to the most complex challenges facing British Columbians.

The DI Program is part of the B.C. government's approach to maximizing the power of data. As a trusted analytics program for public sector data, the program:

- Provides enhanced privacy and security for analysis of government data,
- Streamlines the process for safely linking and analyzing data,
- Enables cross-government interdisciplinary research projects, and
- Boosts government's capacity for data science and advanced analytics.

Government analysts (and eventually academic researchers as well[1]) can use this integrated data to support decision-making on the policies, programs and services that impact the well-being of British Columbians.

## Data has the power to improve lives.

Data is a powerful tool. It can provide deep insights into complex issues, helping government develop better policies and deliver better programs and services to British Columbians.

## About the Privacy and Security Framework

This framework outlines the DI Program's policies and processes that are in place to protect individual privacy.

This framework accompanies the DI Program's Privacy Impact Assessment, which is reviewed and updated every six months. The policies and process articulated in this framework are also subject to ongoing assessment and revision.

In addition to this framework, the DI Program adheres to the B.C. government's Privacy Management and Accountability Policy, as well as the Information Management and Information Technology Management Policy.

Finally, the program is grounded in provincial legislation. Under the *BC Statistics Act,* the program has the legislative authority to link and de-identify data and make it available to project team members, who can then analyze the data within a secure research environment. The Act provides exceptional secrecy protections: data can only be used for statistical purposes, and personal information can't be disclosed from the secure research environment. In addition, the program fully complies with the *Freedom of Information and Protection of Privacy Act.*

[1] Academic users may gain access to the DI Program in Fall 2019.

# Five Safes Model

The Data Innovation Program's privacy and security framework is based on the internationally-recognized Five Safes model – a set of best practices for managing safe access to confidential or sensitive data. Developed by the UK's Office of National Statistics, this model covers five key areas:

1. **Safe Data**
2. **Safe Settings**
3. **Safe Projects**
4. **Safe People**
5. **Safe Outputs**

On their own, each "Safe" adds a layer of protection, reducing the risk of sensitive data being used inappropriately. Together, they form a robust model, embedding strong privacy and security features into the design of the program's technology, data management and business practices.

# Safe Data

## *Data is de-identified*

Within the Data Innovation (DI) Program, only de-identified data is available for analytics. De-identified data is a powerful tool for population-level research projects. It supports analytical insights while maintaining privacy and confidentiality. De-identified data means that:

- Personal identifiers such as names, driver's licence numbers and personal health numbers are removed, but data is protected as personal information; and
- Analysts never see data that identifies anyone personally or can be used to target or make decisions about individuals.

## *Bringing data together*

The DI Program collects data that is voluntarily contributed by data providers (i.e. ministries or other public sector organizations that collect data directly from their clients). Before it can collect this data, the DI Program gives the data provider a Letter of Commitment (LoC), outlining the basis on which the program will collect information, including personal information from the organization, how it will use that information, its approach to privacy and security, and its commitments with respect to the data it is collecting. When an LoC is in place, no Information Sharing Agreement is required.

Once a data provider agrees to contribute data to the program, that data is transferred via a secure upload tool managed by Population Data BC. Once this data has been received, or ingested, it is validated to ensure it matches what the data provider intended to send. This includes ensuring that a dataset's rows, columns, dates and file sizes are accurate.

This data transfer is supported by the DI Program's Data Transfer Protocol, guidelines that instruct data providers on how to prepare the data for transfer. For example, they should encrypt files according to the appropriate B.C. government standards and provide the correct metadata (data about data) as well.

## *De-identification*

Direct identifiers, such as names, phone numbers and personal health numbers, are removed from the content data, and stored and managed separately in an isolated

zone. Only a small subset of program staff – those that link the records – can see direct identifiers. In addition, direct identifiers are never shared with analysts on the project team.

As a further layer of protection, strong *indirect* identifiers are also de-identified. Some of the indirect identifiers that are routinely de-identified include: the date of major life events (e.g. birth, death, marriage, separation), which are truncated to year and month only; and postal codes, which are truncated to the first three characters (i.e. the Forward Sortation Area). This protects individual privacy while maintaining the data's utility.

When identifiers are removed from the content data, a common key is created and retained for both portions. When identity information is linked between datasets, unique, project-specific entity identifiers are created to represent each individual in the data. An association between project-specific entity identifiers and common keys is generated. This association, which is free of identity information, is then used to link de-identified content data.

Data is de-identified in accordance with the DI Program's process (outlined above) and in close consultation with the data providers, who can point out whether any specific variables have additional sensitivities not immediately obvious to the DI Program team.

In all cases, direct identifiers are removed, and strong indirect identifiers are suppressed, truncated or replaced. This approach aims to retain the analytical utility of the data while minimizing the risk of re-identification. Even after the data has been de-identified, it is still protected as personal information and is subject to all privacy and security protections under the *Freedom of Information and Protection of Personal Privacy Act* and the secrecy provisions of the *BC Statistics Act*.

### *Metadata*

To ensure that data is used effectively and responsibly, data providers must provide the DI Program with metadata (data about data).

Metadata is important because it communicates useful details about the data to the DI Program and project team members. Metadata is necessary for severing direct identifiers from content data. It can also help identify any quality issues associated with an incoming dataset. It is also necessary for clearly tracking the personal information that the program collects (similar in principle to a Personal Information Inventory). Finally, metadata records enable the DI Program to maintain a current inventory of all the data in its catalogue. This allows potential project teams to see what datasets are available for analysis before they apply for admittance to the program.

Data providers will be required to provide the DI Program with core metadata, such as core field-level metadata and metadata records. In the future, these records may be released as an additional level of public transparency.

### *Data retention*

Once data has been de-identified, it is held by the DI Program in a research-ready format for approved project teams. The data will remain in the custody and control of the program until it no longer has any analytical value. At that point, data will be decommissioned and no longer made available to new projects. The data and related records would be retained by the Director of Statistics according to the required retention schedule and then destroyed.

Data providers update their data at regular intervals – in most cases annually – to ensure data in the DI Program remains relevant.

### *Transparency*

The DI Program recognizes the importance of being transparent with the public about its activities. Information about the program, including a copy of this privacy and security framework, will be shared on the program's webpage, along with information about the research projects underway. This includes a brief description of the project, the name of the lead ministry and researcher, and a list of the datasets being linked for the project.

# Safe Settings

## *Using the right technology to integrate and store data safely*

Data within the Data Innovation (DI) Program can only be accessed in a secure setting under government's care and control. The program uses a secure research environment located in Canada that:

- Has physical, policy and technological controls to safeguard information;
- Has regular third-party privacy and security reviews and audits; and
- Is managed in partnership by Population Data BC, an academic organization with a 20-year track record without incident.

## *Population Data BC environment*

Population Data BC is a data and education research organization with 20 years of experience in applying the highest standards of physical, technical and procedural privacy protection and data security in providing linking, de-identification and an environment for analytics.

As a service provider, Population Data BC has a General Service Agreement, including a privacy protection schedule, with the DI Program, which ensures they comply with the *Freedom of Information and Protection of Personal Privacy Act*.

Population Data BC manages the DI Program's secure research environment, which prevents unauthorized access and use of data through physical, technological and procedural controls that safeguard data.

A multi-zone virtual and physical environment provides high levels of protection to the most sensitive systems. In addition, all access to services where data is housed is restricted and logged. Furthermore, access to these secure zones requires a physical Yubi-Key, two-factor authentication and a pass phrase.

The DI Program's partnership with Population Data BC includes co-designing technology components to make the current infrastructure more scalable and adapted for a government user-base. The DI Program's security

processes are the same as Population Data BC's security processes.

## *Information Incident Management*

Enquiries, complaints and reports of potential information incidents are managed by the DI Program's Chief Privacy and Security Officer. Both the DI Program and Population Data BC have Information Incident Management policies, which include informing one another in case of an incident or suspected incident involving DI Program data.

The DI Program complies with the B.C. government's Information Incident Management process and other applicable privacy policies and procedures. Any actual or suspected circumstances, incidents or events which jeopardize or could reasonably be expected to harm the privacy of individuals or harm the security of any information system that is used for the program must be reported to the Corporate Information and Records Management Office and the Office of the Chief Information Officer. As appropriate, investigations will be undertaken in coordination with Population Data BC.

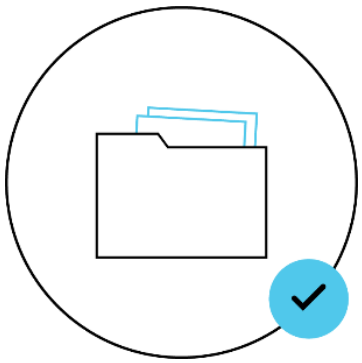
Population Data BC has a strong culture of privacy and security. Along with a dedicated Privacy Lead on staff, all Population Data BC employees participate in annual privacy and security training to ensure these topics are always top of mind, and to keep their processes relevant to emerging trends.

# Safe Projects

### *Projects must be in the public interest*

Only approved projects can access the data. Access will be granted only for projects that:

- Have a clear public benefit;
- Have a valid statistical purpose; and
- Demonstrate sound study design and methodology.



### *Data Access Request*

To initiate a project in the Data Innovation Program, a project team must submit a Data Access Request (DAR), which demonstrates how a proposed project meets the criteria of Safe People and Safe Projects. The DAR must outline its methodology, how the project will benefit the public, and identify the data needed for the project. It must also list all members of the project team, plus their contact information. (Members of a project team only get access to the data requested and approved for use in the DAR.) Finally, the DAR must include publication protocols and standards of conduct for team members.

The DI Program will review the Data Access Request and assess whether it meets all requirements. If it does, the DAR is sent to the Director of Statistics for review and approval.

Currently, only approved government researchers can access the DI Program. Eventually, the program will be open to academic researchers as well, and the DAR includes additional controls for this user group. Namely, academic users must: have ethics approval; pass peer review; disclose third-party involvement (including funding); and demonstrate that any real or perceived conflicts of interest are managed.

Projects must meet different sets of criteria depending on whether they are proposed by government agencies or by academic users[3] (see Figure 1).

Figure 1

**Safe Projects: Requirements for Users**

| | Demonstrate public benefit | Request de-identified data relevant to project objectives | Support gov't goal, ministry mandate or program/service improvement | Have ministry approval and commitment of resources | Have ethics approval, and pass peer review | Meet acceptable third-party involvement or funding sources |
|---|---|---|---|---|---|---|
| Government users | | | | | | |
| Academic users | | | | | | |

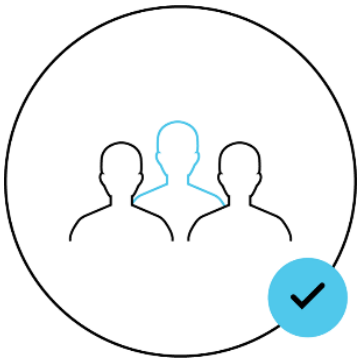
# Safe People

*Only authorized individuals can access the data*

Only authorized people will ever receive access to data within the Data Innovation (DI) Program. These authorised users include Population Data BC staff, government employees, and eventually academic users.

Population Data BC staff with access to DI Program data are required to take the secrecy oath under the *BC Statistics Act* and adhere to the privacy and security conditions of the General Service Agreement with the DI Program. A limited number of Population Data BC staff have access to direct identifiers for the purpose of linking data. These people must also sign a confidentiality agreement, complete a criminal record check, and take specialized initial and annual privacy training. In addition, they are subject to audits.

Generally, Safe People refers to people who are gaining access to de-identified data as members of an approved project team (see Safe Projects). Project team members can be trained government analysts, government-contracted researchers, and eventually academic users.



Project team members are required to:

- Sign Engagement Agreements that stipulate terms and conditions of data access and use,
- Take an oath of secrecy under the *BC Statistics Act*, and
- Complete the DI Program's training and pass an exam.

Project team members may need to meet additional criteria depending on whether they are a government employee or part of an external project team, as shown in Figure 2.

Figure 2

**Safe People: Requirements for Users**

| | Engagement Agreement | *BC Statistics Act* Oath of Secrecy | Merit-based hiring, Oath of Employment, Standards of Conduct | History of safe data use | Any conflicts of interest are successfully managed | Academic supervisor must support student applications |
|---|---|---|---|---|---|---|
| Government users | | | | | | |
| Academic users | | | | | | |

### *Engagement Agreement and oath of secrecy*

Each project team member must sign an Engagement Agreement, obliging them to comply with the security and secrecy requirements of the program. The Engagement Agreement is the tool by which individual project team members are deemed temporary employees under the *BC Statistics Act*. It includes terms and conditions related to access and use of data and use of the DI Program's secure research environment managed by Population Data BC.

In addition, before members of a project team can access any de-identified data, they must be engaged by the Director of Statistics and take the *BC Statistics Act* oath of secrecy. This oath obliges them not to disclose any information that comes to their knowledge through their work under the *BC Statistics Act*.

### *Mandatory training for project team members*

All project team members are required to complete the DI Program's specific training on privacy and statistical disclosure controls. The first training module covers basic privacy concepts, how to use data in the secure research environment, and the process for managing a breach or suspected breach of privacy. The second module thoroughly covers the program's output requirements, including significant instruction on statistical disclosure controls.

Statistical disclosure controls are the methods used to ensure research results can't identify individuals, even accidentally. These methods, which are integral to Safe Outputs (see below), include suppression, aggregation, thresholds (sometimes called minimum cell size), and random rounding, among others.

## Safe Outputs

### *Additional protection of privacy in research outputs*

The Data Innovation Program takes measures to ensure project results are always anonymous and non-disclosive (meaning they can't identify individuals). These include:

- Setting clear obligations under the terms and conditions of access, and
- Ensuring project results are fully anonymized through both technological tools and a manual review by a statistician.

### *Anonymization and accuracy*

Under the *BC Statistics Act*, all data must be anonymized before it can be released from the secure research environment managed by Population Data BC. The secrecy of information collected by the DI Program is an essential aspect of the program. So, before outputs can be released, they must be de-identified to the point where an individual cannot be re-identified. First, they

are submitted to an automated scan that checks for hidden data, as well as statistical disclosure controls applied by the user, who is given resources to help with this step. Then, the outputs are manually reviewed by a statistician prior to release.

Finally, prior to publication, outputs must be submitted to DI Program staff, who will share them with the original data providers to check the material has been appropriately referenced and there has been no gross mischaracterisation of the data.

BRITISH COLUMBIA | Ministry of Citizens' Services

# For more information

For more information about privacy and security within the Data Innovation Program, please contact:

**Beth Collins**
**Chief Privacy and Security Officer**
Digital Platforms and Data Division
Office of the Chief Information Officer
Ministry of Citizens' Services

**(250) 217-2108**
**Beth.Collins@gov.bc.ca**