Leveraging data for impact analysis: Module 4

Shelley Jeglic, Data Access Division

March 21, 2024

Delivering insight through data for a better Canada
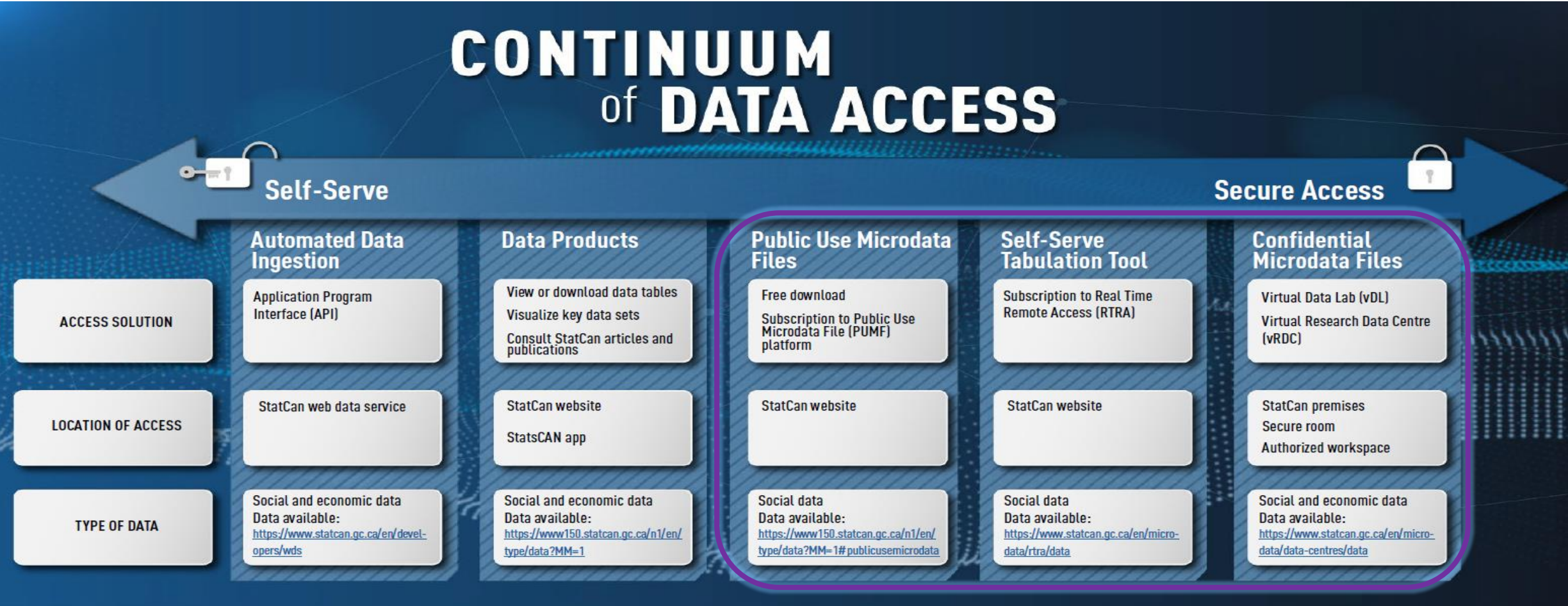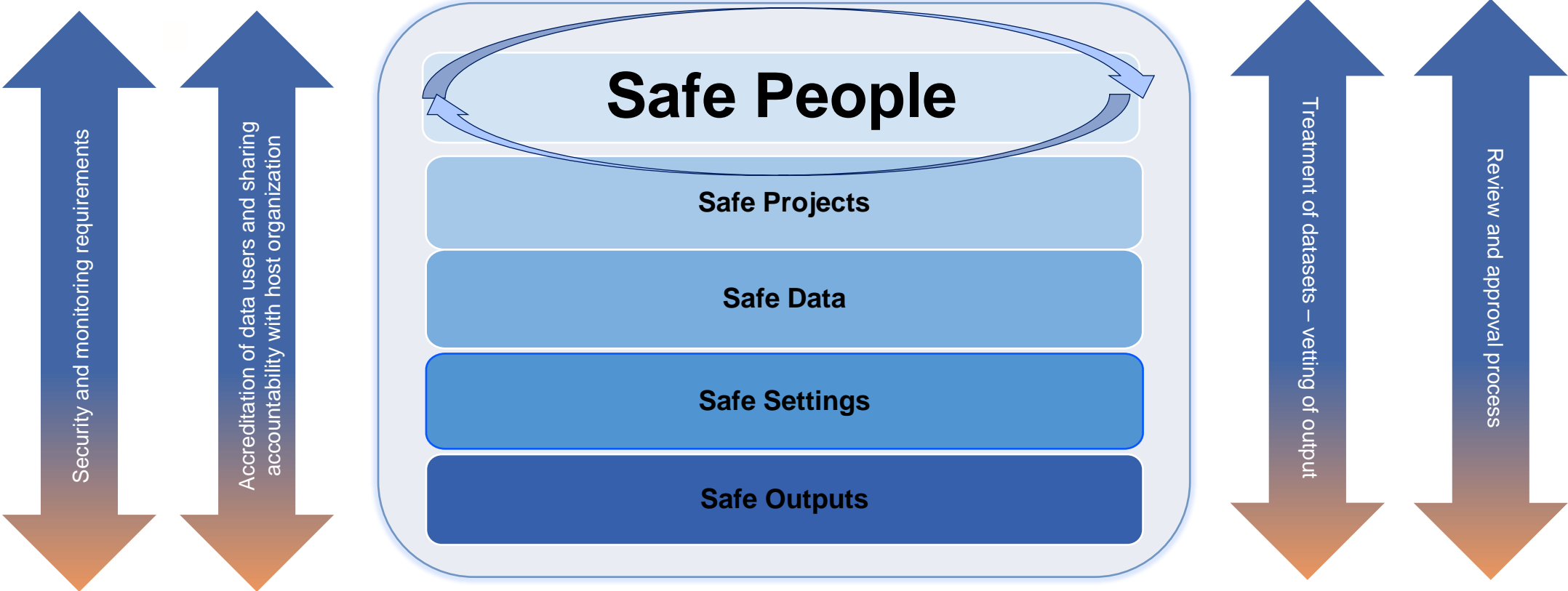
Statistics Canada Data Access Services

What we offer: Accessible representational meaningful data about Canadians for Canadians

CONTINUUM of DATA ACCESS

Self-Serve ⟷ Secure Access

	Automated Data Ingestion	Data Products	Public Use Microdata Files	Self-Serve Tabulation Tool	Confidential Microdata Files
ACCESS SOLUTION	Application Program Interface (API)	View or download data tables; Visualize key data sets; Consult StatCan articles and publications	Free download; Subscription to Public Use Microdata File (PUMF) platform	Subscription to Real Time Remote Access (RTRA)	Virtual Data Lab (vDL); Virtual Research Data Centre (vRDC)
LOCATION OF ACCESS	StatCan web data service	StatCan website; StatsCAN app	StatCan website	StatCan website	StatCan premises; Secure room; Authorized workspace
TYPE OF DATA	Social and economic data Data available: https://www.statcan.gc.ca/en/developers/wds	Social and economic data Data available: https://www150.statcan.gc.ca/n1/en/type/data?MM=1	Social data Data available: https://www150.statcan.gc.ca/n1/en/type/data?MM=1#publicusemicrodata	Social data Data available: https://www.statcan.gc.ca/en/microdata/rtra/data	Social and economic data Data available: https://www.statcan.gc.ca/en/microdata/data-centres/data

Using the Five Safes Framework as Guiding Principles

Security and monitoring requirements

Accreditation of data users and sharing accountability with host organization

Safe People

Safe Projects

Safe Data

Safe Settings

Safe Outputs

Treatment of datasets – vetting of output

Review and approval process

The 5 Safes are considered in combination - and on a sliding scale - depending on type of researcher, access and data.

Statistics Canada   Statistique Canada

Canada

Public Use Microdata Files

Access

- Subscribers of the **Public Use Microdata Files (PUMFs) Collection** are granted access to all anonymized and non-aggregated data in the PUMF collection

Ideal for

- Data Exploration, Extractions and as an Analytical tool for Academic and Policy Research:
  - Calculating frequencies, cross tabulations, means, percentiles, percent distribution, proportions, ratios, and shares
- Use as an analytical training tool

Benefits

- In-depth searches with selected and rounded outputs in file format
- Operates on an open license
- Principal product used for teaching data literacy
- Data are all cautiously modified and reviewed for research purposes

Fees: PUMF Collection

- Yearly institutional access fee (access allowed through IP range recognition)
- **$10,000 per year**

PUMF Access

All new PUMFs can now be downloaded for free from the StatCan website.

- Older PUMFs will be made available in phases

PUMF Collection program

- The subscription is available to government departments and businesses from inside and outside Canada.
- Unlimited access to all microdata and documentation is available through StatCan's Electronic File Transfer Service (EFT).
- Select files are available through Rich data Services (RDS) analytical platform.

Data Liberation Initiative (DLI) program

- The membership is available to post-secondary institutions.
- Unlimited access to all microdata and documentation is available through StatCan's Electronic File Transfer Service (EFT).
- Select files are available through Rich Data Services (RDS) analytical platform.

PUMF Access: Rich Data Services (RDS)

Features
- [RDS Landing (statcan.gc.ca)](statcan.gc.ca)
- IP restricted self-serve online database
- Bilingual and accessible user friendly web interface
- Files downloadable in different formats
- Application Program Interface (API) compatible
- Data Documentation Initiative (DDI) compliant

RDS Explorer
- Used to browse the data records, create custom extracts by filtering records and selecting variables, and produce personalized open data packages for download.

RDS Tabulation Engine
- Used to aggregate data across various dimensions to rapidly create analytical tables.

Real Time Remote Access

Access

- The **Real Time Remote Access (RTRA)** tool is an online tabulation tool enabling quick access to microdata through SAS programs.

- It is a subscription-based service.

Benefits

- Conduct analysis on descriptive statistics

- Quick access to releasable tabulations

- RTRA **does not** gain users direct access to the microdata, nor does it allow users view microdata file content.

Ideal for

- Generating a full range of descriptive statistics that can be used for research, training and policy briefings

- Calculating frequencies, means, percentiles, proportions, ratios and share

Fees

- **$10,000/year (up to 10 user accounts)**
- **Additional user accounts $2,000/account**

Confidential Microdata Files

Access

- Access through Authorized Workspaces using Cloud technology via office or personal residence
  - Virtual Research Data Centre (vRDC) for academics
  - Virtual Data Lab (vDL) for government and other organizations
- Includes Research Data Centre and Federal Research Data Centre, and Secure Rooms

Benefits

- Expert support and direct access to detailed confidential microdata files
- Secure 24/7 remote access to anonymized data under certain criteria available to researchers through secure virtual IT infrastructure
- Greater opportunities for collaboration

Ideal for

- Answering research questions that require the use of advanced analytical methods such as complex multivariate analysis; modelling

Fees

- Fees for access vary depending on project parameters

Statistics Canada   Statistique Canada

Canada

Microdata Access: Virtual Data Lab

- The Virtual Data Lab (vDL) is a secure cloud infrastructure used to store and access microdata research projects.

- The vDL provides a flexible approach to accessing Statistics Canada microdata, and includes monitoring controls and built-in vetting management.

- Through "**authorized workspaces**" for eligible projects, qualifying researchers can access microdata projects while teleworking from home or their own office.

- The vDL replaced the need for an on-site secure room and/or StatCan issued laptops for data access.
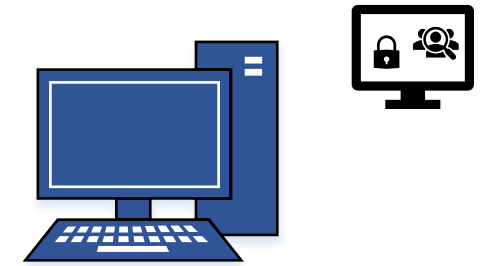
Authorized Workspace
A closed office or conference room within the premises of an organization, or a private space within the Deemed Employee's personal residence.
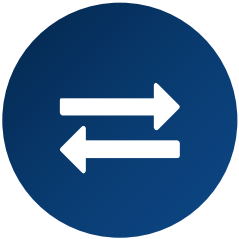
Highlights and Features of vDL

Secure 24/7 remote access

Robust IT security measures

Shared-risk approach

Housed in Collaborative Analytics Environment

Secure importing/exporting from environment

Two-factor authentication

Monitoring and reporting capabilities

Replaces requirement for secure room

Accessing Microdata: Submitting New Microdata Project Proposals

Applications for microdata access are submitted through the Microdata Access Portal (MAP) https://www.statcan.gc.ca/rdc-cdr/eng/user/login

The process for submitting a new application includes:

- Creating a MAP Account
- Determining the Type of Project & Providing Project Details (including the project proposal following project proposal template guidelines)
- Completing Co-Investigator Profiles
- Application Submission

For assistance in completing an application, please contact: statcan.dad-apu-dad-uta.statcan@statcan.gc.ca

Organization Requirements for VDL Access

Eligible Organizations

- Must have a signed Agreement
  - Section 10 agreement for Provincial Ministries
- Must be an accredited (known and trusted) organization
  - Complete an Organization Accreditation Documentation
- Must appoint a Data Access Liaison

Researcher Requirements for vDL Access

Eligible Researchers

- Become a Deemed Employee of Statistics Canada
  - Swear or Affirm the Oath of Office and Secrecy
  - Obtain reliability security clearance
- Complete mandatory training and must pass mandatory test
- Must be accredited(affiliation to an accredited organization increase the level)
- Data Access does not pose a Conflict of Interest
- An approved Microdata Research Project Proposal
- Signed Data Access Microdata Research Contract with StatCan

Costs for Virtual Data Lab

Billing
Billing is completed by Purchase Confirmation (less than 20K per) or by Letter of Agreement (LOA) (20K or greater).

Costs
The Federal Research Data Centre (FRDC) operates on a cost-recovery basis. All fees are inclusive of project set-up, account creation, cloud costs for project hosting, analyst support and vetting output for 12 months of continuous access.

<u>**Base Project: $13,500**</u>
- 1 to 3 researchers accessing from an "authorized workspace" in the virtual Data Lab (vDL) or onsite in the Federal Research Data Centre (FRDC)
- 12 months of continuous access
- A maximum of 1,440 access hours (across vDL Virtual Machines (VM) or onsite researchers)
- Standard Software Suite
  - Project Libre (open-source software equivalent of the MS Office suite)
  - STATA MP64
  - IDLE (Python 3.9)
  - Python 3.9
  - SAS Enterprise Guide 8.1
  - R 64 4.1.0
  - IBM SPSS Statistics 24
  - Libre Office
  - R Studio 1.4.1717
  - SAS 9.4 (Fr+En)
  - Miniforge3 (to replace Anaconda in near future)

<u>**Project Extension: $13,500 per year or $1,125 per month**</u>
- Applies only to projects over 12 months
- 1 to 3 vDL Researchers or FRDC On-Site Researchers
- A maximum of 120 access hours per month (across vDL Virtual Machines (VM))
- Standard Software Suite
- Additional researchers may also be extended at a cost per researcher of $2,500 (12 months) or $208 (per month)

Note: Project extensions of less than 12 months must end or close out at the end of the extension period or additional fees may be incurred for multiple extensions in the same fiscal year.

<u>**Project Add-Ons:**</u>
- $2,500 per additional researcher with on-site FRDC access or vDL access (12 months of access)
- $500 for VM researcher replacement/change (for projects at their maximum number of researchers)
- $500 per project participant (researchers without a VM) for projects with 3 or more VMs
- Non-standard software - based on software cost
- Additional computing power- based on computing requirements
- Additional services/features as required

<u>**Fees effective April 1, 2023. Applicable taxes are extra.**</u>

Protecting Data Confidentiality

Confidentiality Vetting

The process of reviewing a statistical output to ensure it does not pose a disclosure risk to data confidentiality

Vetting Rules

Set of strategies to minimise disclosure risks. E.g.,

*Minimum contributing observations

*Application of survey weights

*Rounding counts and $ values

*Limits to low-levels of geography

*Tests of dominance

Vetting Timelines

Many factors to consider:

*Complete and accurate supporting documentation

*Size and complexity of request

*Interim verses final version

*Previous output requests

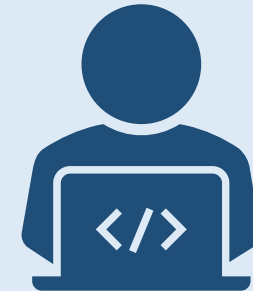Statistics Canada | Statistique Canada

Canada

Treasury Board of Canada Secretariat - StatCan Data Access Liaison

Treasury Board of Canada Secretariat's Data Access Liaison is Statistics Canada's primary contact for communicating information related departmental microdata access.

Diana Macecek
Diana.macecek@tbs-sct.gc.ca

Contact Information

Public Use Microdata Files

PUMF collection: https://www150.statcan.gc.ca/n1/en/type/data?MM=1#publicusemicrodata

Inquiries or assistance: statcan.fmgd-pumf.statcan@canada.ca

Real Time Remote Access

List of available data**:** https://www.statcan.gc.ca/en/microdata/rtra/data

Inquiries or assistance: statcan.dad-rtra-dad-adtr.statcan@statcan.gc.ca

Research Data Centres

List of available data: https://www.statcan.gc.ca/en/microdata/data-centres/data

Inquiries or assistance: statcan.dad-apu-dad-uta.statcan@statcan.gc.ca

Thank You

Shelley Jeglic
Assistant Director
Data Access Division
shelley.jeglic@statcan.gc.ca

Geneviève Jourdain
Director
Data Access Division
genevieve.jourdain@statcan.gc.ca