



Statistics
Canada

Statistique
Canada

Access to Statistics Canada's Microdata

www.statcan.gc.ca



Telling Canada's
story in numbers

September 21, 2017

Canada 



Overview:

- Legislative Framework
- Users of Microdata
- Continuum of Access Framework
- Access to social microdata
- Access to business and economic microdata
- Modernization of microdata access at Statistics Canada



Legislative Framework:

- In Canada, providing national statistics is a federal responsibility falling under the Statistics Act, the Privacy Act, the Income Tax Act, and The Excise Tax Act
- The Statistics Act requires STC to collect, compile, analyze and publish statistical information...*all while*
- Maintaining confidentiality of Canadians' information
- Access programs have been designed to balance users' needs with confidentiality protection measures



Micro-Data Users:

- University professors and students need access to microdata for courses and assignments
- Government need access to microdata that can be accessed quickly, analyzed and disseminated without worry of disclosure
- Government and academics researchers need access to more detailed microdata for their complex analyses

Continuum of Access:





Public Use Microdata Files (PUMF):

- PUMFs contain microdata which have been modified to ensure that no one individual can be identified from a combination of variables on the file
- The primary users are universities, local governments and small businesses.
- The file is anonymized utilizing a number of methods including:
 - Suppression of identifying variables
 - Regrouping of some variables into grouped categorical variables
 - Data suppression where regrouping is not sufficient
 - Reduction in the geographic detail
- The major limitation of a PUMF is its restricted analytic potential



Real Time Remote Access (RTRA):

- RTRA is:
 - On-line remote program using SAS
 - Users do not view data
 - A full range of descriptive statistics is available through the RTRA such as frequencies, means, medians, percentiles, proportions and ratios.
 - Results available in in real-time
- Users are from government, non-profit organizations and academic institutions



Protecting confidentiality in RTRA

- Restrict
 - Available SAS codes prevent listing of data.
 - Number of submissions per day
- Sensitive variables are removed: specifically sensitive and low level geographical variables
- Rounding applied
 - Controlled rounding is applied independently on each cell, including subtotals
 - Totals are additive
- Only weighted analyses can be produced
- Minimum threshold suppression rules do not apply to RTRA



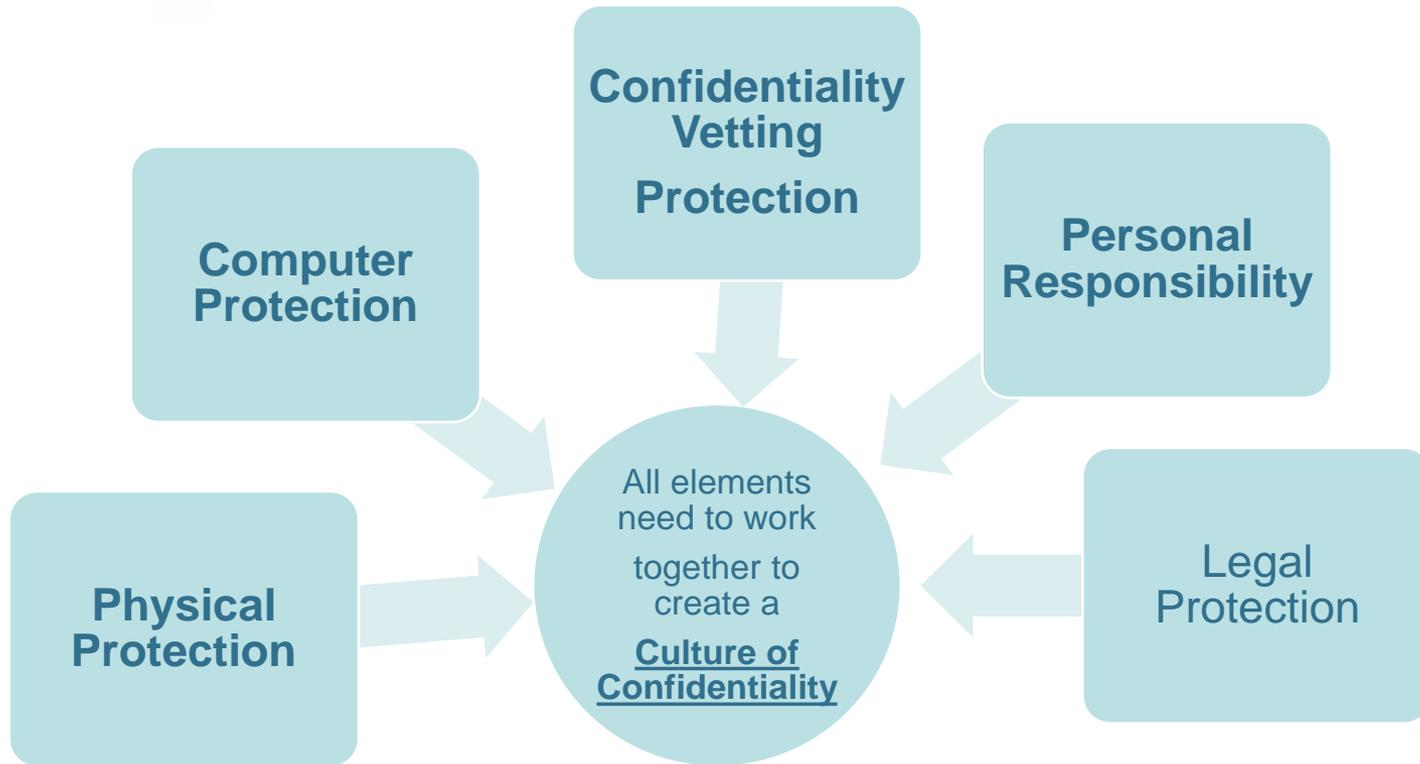
Real Time Remote Access

Limitations of the RTRA program

- Restriction to SAS
- Removal of sensitive variables, especially geography
- Limited descriptive statistics
- Lack of significance testing
- No modeling capacity



Research Data Centres (RDCs)



Inspections and audits are conducted regularly to ensure adherence to the security procedures



Protecting confidentiality in the RDCs

- Data are de-identified
- All output reviewed by Statistics Canada's analysts before release
- Confidentiality Vetting Strategies:
 - Minimum cell sizes based on sensitivity of the variables and the size of the sample
 - Weighted results are releasable
 - No low-level geography are releasable
 - Homogeneity and dominance tests for income – Tax and census data
 - For administrative data sets, controlled rounding is used and a scoring method is utilized to assess the risk of disclosure
- All released output are emailed to the researcher once vetted for confidentiality



Research Data Centres (RDCs)

Limitations

- Access times not flexible: Security protocols require staff to be present at all time during operating hours.
- Researchers view confidentiality vetting as:
 - Burdensome
 - Confidentiality rules are too complex and restrictive
 - Results in delays
- Data collection limited to social data



Centre for Data Development and Economic Research (CDER)

- CDER provides access to business and economic data at Statistics Canada's Head Office
- Users of CDER include researchers from: Canadian and foreign institutions, think tanks, and government departments
- CDER users view only their aggregate and analytical results but not the actual microdata
- Some files have synthetic versions for preliminary analysis.
- Firm-level data differs from that of the individual-level data
 - the distribution of firm-level variables are often skewed
 - the data are sparse in certain dimensions (e.g., industry and geography)
 - certain firms dominate their industries



Protecting Confidentiality: Firm-level Data

Risk mitigation strategies for confidentiality vetting

- Limit release of tabular output to only what is necessary for publication
- Centralized vetting in CDER consultation with subject-matter data producers and experts
- Automated in-house generalized system for disclosure analysis
 - Utilizes dominance and homogeneity rules, suppresses for minimum thresholds
 - Some functions to deal with residual disclosure (it is possible to tell the system that certain cells have been released)
 - No rounding, but applies complementary suppression

Limitations of CDER

- CDER faces similar challenges as the RDC program
 - Access to business microdata only in Statistics Canada's headquarters



Modernization of Microdata Access

The Pillars of Modernization

User Centric	Users have the information and data they need, when they need it, in the ways they want to access it, with the tools and knowledge to make full use of it
Sharing & Collaboration	To continually develop and enhance meaningful collaborative partnerships that expand the reach and use of statistical information, both within Statistics Canada and with our partner organisations
Leading Edge Methods & Data Integration	Access to new or untapped data; modify the role of surveys; greater reliance on modelling and integration capacity through R&D environment
Statistical Capacity Building & Leadership	To be leaders in identifying, building and fostering savvy information and critical analysis skills beyond our own perimeters



For further information contact:

Donna.Dosman@Canada.ca

<http://www.statcan.gc.ca/eng/start>