



Statistics
Canada

Statistique
Canada

Accessing Statistics Canada Health Data at McMaster

www.statcan.gc.ca



CANADA 150

Telling Canada's
story in numbers

**Accessing Statistics Canada Health
Data at McMaster
October 16, 2018**

Peter Kitchen

rdc@mcmaster.ca

Vivek Jadon

vivek@mcmaster.ca

Canada 

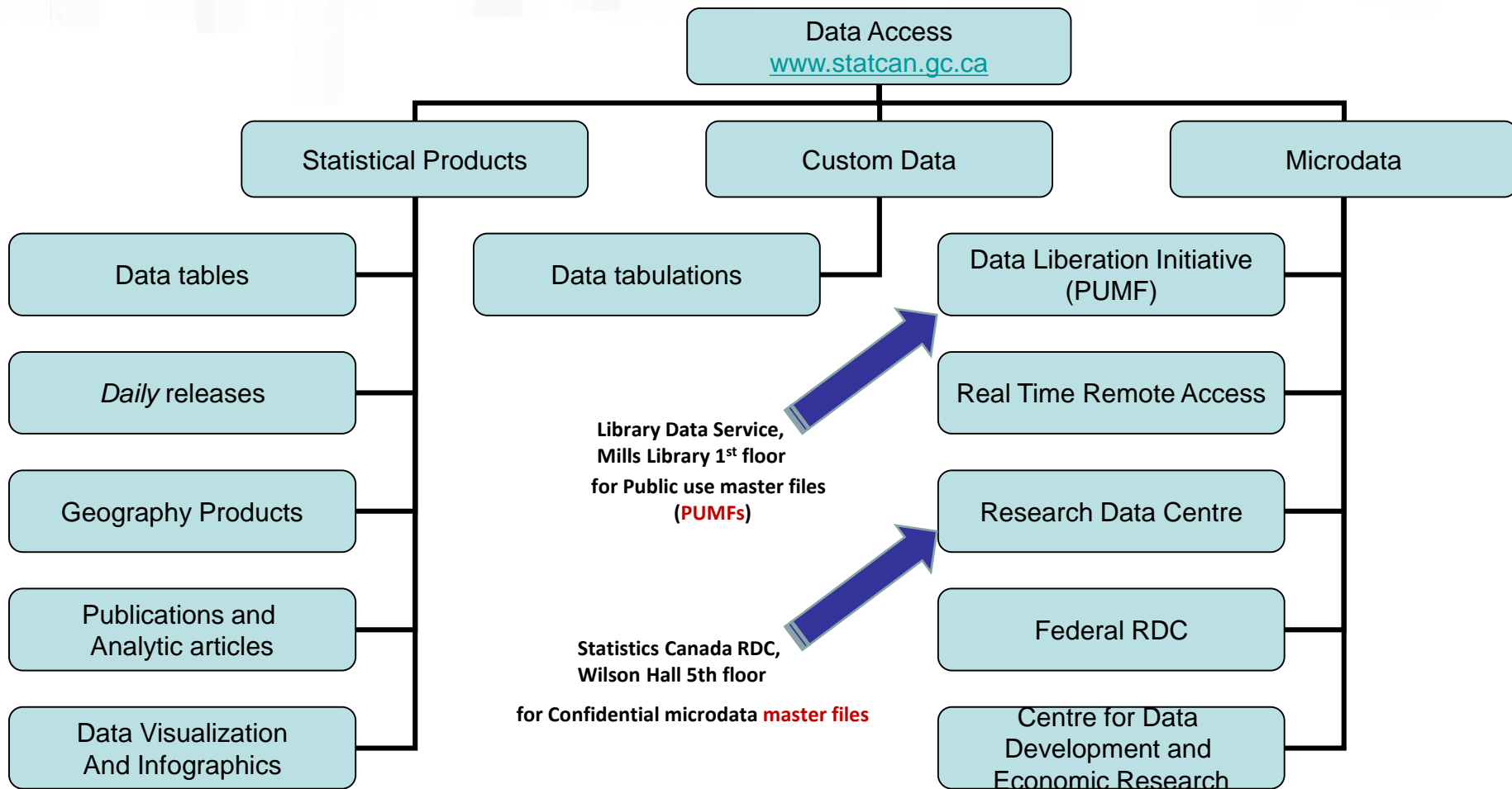


Session Outline

- Continuum of Data Access @ Statistics Canada
- Microdata Access Programs @ Statistics Canada
 - Data Liberation Initiative (DLI)
 - Real Time Remote Access (RTRA)
 - Research Data Centre (RDC)
- Canadian Institute for Health Information (CIHI)
- Live demos
 - Odesi Data Portal and Stat Can websites
 - RDC and CRDCN websites
- Q & A/Discussion



Continuum of Data Access





CONTINUUM of MICRODATA ACCESS

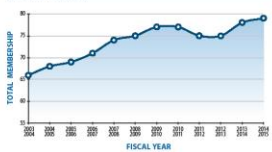


Nonrestricted Access

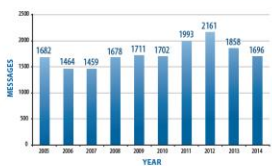
DATA LIBERATION INITIATIVE (DLI)

- the DLI is a partnership that started in 1996 between post-secondary institutions and Statistics Canada, formalized under a licence agreement for improving access to Canadian data resources
- Canadian post-secondary institutions can access over 350 standard products, such as Public Use Microdata Files (PUMFs), on machine-readable files
- modifications performed by Statistics Canada, including collapsing, suppressing, and removing variables before PUMFs are released, ensure that the risk of breaching confidentiality has been removed
- support and training for each DLI contact enable contacts to train researchers / professors / students and employees on how to use information from the DLI collection
- a listserv, an electronic mailing list, exists where questions related to the collection can be asked and answered by Statistics Canada, the DLI community or specialists
- data accessed through the DLI are used for research, training and planning purposes

DLI MEMBERSHIP



DLI LIST MESSAGES BY YEAR



ACCESS TO PUBLIC USE MICRODATA FILES (PUMFs) COLLECTION

- access for Canadian or international organizations to standard products on machine-readable files under a licence agreement
- support to researchers to enable them to use the information from the PUMF collection
- a listserv, an electronic mailing list, exists where questions related to the collection can be asked and answered by Statistics Canada or specialists
- data accessed through PUMFs Collection are used for research

REAL TIME REMOTE ACCESS (RTRA)

- online remote access facility allowing users to submit programs to run, more or less in real time, data analyses on microdata located in a central, secure location
- Statistical Analysis Software (SAS) programs that run descriptive statistics
- confidentiality applied automatically
- no primary manual interventions
- currently over 60 data files
- ability to interrogate RTRA microdata at any time
- access from any computer with Internet access, using a secure username and password – no travel to RDCs required

RESEARCH DATA CENTRES (RDC)

- a partnership with universities, Statistics Canada, and Canadian research funding agencies to provide access to researchers, in a secure university setting, to microdata from population and household surveys and administrative data
- currently over 350 confidential microdata files
- university centres are accessible to academic, government and other researchers
- one RDC in Ottawa provides access to these data for federal government researchers
- purpose of access is to conduct complex statistical analysis (these data can be longitudinal, survey, administrative or linked, and have detailed geographies)
- access granted to researchers with approved projects and security clearance
- provide support and training to RDC researchers in the areas of statistical methods and data

CANADIAN CENTRE FOR DATA DEVELOPMENT AND ECONOMIC RESEARCH (CDER)

- repository for business and economic microdata files containing sufficient detail for researchers to undertake meaningful and complex analysis while safeguarding the confidentiality of the results and the privacy of business respondents and workers in these establishments
- two-stage review-process feasibility and detailed project review/costing
- extracts of microdata are provided based on approved project requests
- access granted to researchers with approved projects and security clearance
- CDER works with researchers and Statistics Canada to produce specific analytical data files

Restricted Access

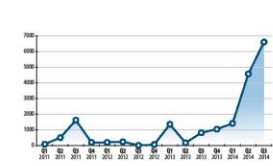


ACCESS IS A KEY PILLAR IN STATISTICS CANADA'S CORPORATE BUSINESS PLAN:

"Increasingly, modern policy research requires access not only to aggregated statistics but also to data at the level of the individual business, household or person. Statistics Canada is committed to removing obstacles to access while preserving the privacy and confidentiality of respondents."

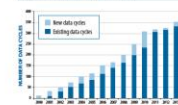
(CORPORATE BUSINESS PLAN 2014-15 TO 2016-17, STATISTICS CANADA)

AVERAGE NUMBER OF SUCCESSFUL SUBMISSIONS

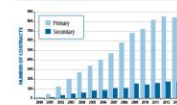


FOR MORE INFORMATION:
www.statcan.gc.ca

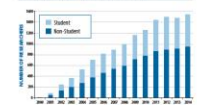
NUMBER OF DATA CYCLES IN THE RDCs



AVERAGE MONTHLY ACTIVE CONTRACTS



AVERAGE MONTHLY COUNTS OF RESEARCHERS



LOCATION OF RDCs ACROSS CANADA



SECURITY REQUIREMENTS FOR AN RDC





Statistics Canada @ McMaster

- **Online access:** www.statcan.gc.ca

- **In-person access:**
 - Library Data Service: Mills Memorial Library, 1st Floor
 - Research Data Centre: L.R. Wilson, 5th Floor

- **Remote access:** Real Time Remote Access (RTRA)



Microdata Access Programs

- “Public Use” Microdata access:
 - Data Liberation Initiative (DLI)

- “Restricted” Microdata access:
 - Real Time Remote Access (RTRA)
 - Research Data Centres (RDC)
 - Federal Research Data Centre (FRDC)
 - Canadian Centre for Data Development and Economic Research (CDER)



Data Liberation Initiative (DLI)

- DLI provides access to Stat Can's Standard products, Databases, 350 Public use Microdata sets and Geographic information files.
 - Main focus of DLI collection is on Socio-Economic data: Health, Education/Literacy, Labor Market, Income, Travel, Justice, Census of Population etc.
 - Databases such as the Small Area Business and Labour Database, Inter-Corporate Ownership, Financial Performance Indicators, Trade data etc.
 - An enhanced line of Census products
 - Aggregated data on subject such as Justice and Education
 - All standard geographic files and databases including PCCF and PCCF+



Data Liberation Initiative (DLI)

- Metadata from DLI surveys are marked-up in DDI/XML format for discoverability at the variable level from Statistics Canada's **Nesstar site** and **Odesi Data Portal**.
- DLI members have support through a very active listserv.
- Currently 77 subscribing institutions -- McMaster University Library is part of Stat Can DLI program.



Microdata Access

- Public Use Microdata File (PUMF)

- Each Public Use Microdata File is based on a corresponding master data file. The modifications performed by Statistics Canada before the PUMF is released ensure that the risk of breaching confidentiality has been removed. Since the results of any analysis performed do not have to be scrutinized before they are released, the file is considered “Public”.
- Modifications made to the Master files for conversion to PUMFs may include: **collapsing of variables** (e.g., age groups instead of individual years of age); **collapsing variables into one variable** (e.g., multiple language questions collapsed into one language variable for analysis); **suppressing variables** (although the variable is part of the master file, it will not show up in the public file); and **removing outliers** (removing cases that are extremes - often used with income).
- By using these techniques to anonymise the files, combining variables will not result in the user identifying a respondent from any given survey.



Microdata Access

- Public Use Microdata File (PUMF)
 - Benefits
 - Free
 - Very few restrictions on access & use of the data
 - No approval process to access the data
 - Limitations
 - Content is limited (screened and grouped for confidentiality)
 - Not all surveys have a PUMF
 - PUMFs are cross-sectional, i.e., represent data collected at one point in time



Microdata Access: Master File vs PUMFs

Compared with PUMFs, a *master file*:

- contains the full sample of respondents, not a sub-set of them
- includes additional categories in variables; more detailed information
- allows research on lower levels of geography
- provides discrete values for certain variables, such as age or body weight, instead of categories (as found in PUMFs)
- may offer other concepts that are not available in PUMFs

Moreover, master files contain derived variables and bootstrap weights

- RDC access is useful when PUMF does not exist or provide adequate level of details for quality research or when longitudinal data linkage is required



Microdata Access

■ Synthetic Files

- These microdata do not contain actual “real” cases but contain “**pseudo-cases**” that provide aggregate results close to the “real” cases
- These files have been prepared to create analysis runs with the master file without possibly disclosing or identifying any of the cases
- The results are NOT to be reported, but are strictly to be used to prepare analysis of master files
- Usually associated with longitudinal files, e.g. NLSCY, NPHS etc.



Real Time Remote Access (RTRA)

- On-line remote access facility allowing users to run SAS programs, in more or less in real-time, against microdata located in a central and secure location.
- Researchers using the RTRA system do not gain direct access to the microdata and cannot view the content of the microdata. Instead, users submit SAS programs to extract results in the form of frequency tables.
- As RTRA researchers cannot view the microdata, becoming a deemed employees of Stat Can is no longer necessary. Hence, rapid access to “confidential” microdata files.
- Using a secure username and password, the RTRA provides around the clock access to survey results through EFT from any computer with internet access.
- Confidentiality of the micro data is automated in the RTRA system, eliminating the need for manual intervention and allowing for rapid access to results.



Real Time Remote Access (RTRA)

■ Benefits:

- Access from any computer with internet access using a secure username and password - No travel to RDCs
- Few conditions on access and use of data
- Full Master Files available; Confidentiality automated
- Deemed employee status not required

■ Limitations:

- Descriptive statistics only (**Frequency, Mean, Median, Percentiles, Percentage distribution, Ratios, Proportions**)
- Can only be used with SAS
- Not all data sets available under RTRA service – currently 105 data sets (no access to Census Microdata)
- Costs – 2 accounts for \$5,000; unlimited access for \$10,000

- <https://www.statcan.gc.ca/eng/rtra/rtra>

Health PUMF's - DLI

- Canadian Community Health Survey (CCHS): 2001 – 2014 and Synthetic data from 2009 - 2012
- National Population Health Survey (NPHS): 1994/95, 1996/97, 1998/99 and Synthetic data from 1994/95 – 2000/01
- Canadian Tobacco Use Monitoring Survey (CTUMS): 1999 – 2012
- Canadian Alcohol and Drug Use Monitoring Survey (CADUMS): 2008 – 2012
- Canadian Tobacco, Alcohol and Drugs Survey (CTADS): 2013 and 2015
- Canadians' Attitudes Toward the Health Care System 2011
- Discharge Abstract Database (DAD) – Hospital Discharge Data: 2009/10 – 2014/15
- HIV/AIDS Attitudinal Tracking Survey 2012
- Youth Smoking Survey (YSS): 1994 – 2010/11

Health Statistics

- Statistics Canada's website
 - Access Data Tables, Thematic Maps, Data Visualizations
 - Browse Information such as Data, Analysis and Reference on specific subject (Health)
 - Access Health related Analysis such as Stats in brief, Articles/Reports and Journals/Periodicals
- Library Data Service (LDS) Data/Statistics Collection
 - Access to a variety of resources on Health classified into Open Data, Statistics, Microdata and Data Visualizations

Geography Resources on Health

- [Postal Code Conversion File \(PCCF/PCCF+\)](#): links six-character postal codes to standard geographic areas such as DAs, Census tracts and Census subdivisions. **Available through LDS.**
- [Health Regions: Boundaries and Correspondence with Census Geography](#)

Canadian Institute for Health Information (CIHI)

- Independent not-for-profit organization, EST. 1994
- Collects and provides information and data on Canada's health system & the health of Canadians
- Partners include Statistics Canada, Health Canada, as well as Health Ministries across the country

Canadian Institute for Health Information (CIHI)

- CIHI's data holdings fall under three categories:
 - Health personnel
 - Health Spending
 - Health Services

CIHI's Data Holdings

■ Health personnel

- National Physician Database (NPDB)
- Registered Nurses Database (RNDB)
- Pharmacist Database (PDB)
- Occupational Therapist Database (OTDB)
- Medical Radiation Technologist Database (MRTDB)
- Medical Laboratory Technologist Database (MLTDB)
- Physiotherapist Database (PTDB)
- Scott's Medical Database (SMD)

CIHI's Data Holdings

- **Health spending**

- Canadian Management Information System Database (CMDB)
- National Health Expenditure Database (NHEX)

CIHI's Data Holdings

■ Health services

- Discharge Abstract Database (DAD)
- National Prescription Drug Utilization Information System (NPDUIS)
- Organization for Economic Co-operation and Development (OECD) Health Database (Canadian Segment)
- National Ambulatory Care Reporting System (NACRS)
- Canadian Joint Replacement Registry (CJRR)
- National Trauma Registry (NTR)
- Ontario Mental Health Reporting System (OMHRS)
- Hospital Morbidity Database (HMDB)
- Canadian Organ Replacement Register (CORR)
- Continuing Care Reporting System (CCRS)
- Hospital Mental Health Database (HMHDB)
- Canadian Multiple Sclerosis Monitoring System (CMSMS)

CIHI's GSDAP Program

- CIHI provides data to qualifying graduate students at no cost through the Graduate Student Data Access Program (GSDAP).
- Graduate students conducting research can access data through various databases and clinical registries in the following areas:
 - Health Services
 - Health Spending
 - Health Human Resources
- For more information, please visit [GSDAP program website](#).



Odesi Data Portal

- **Odesi is...**

- a web-based data extraction system delivered through Scholars Portal, provides access to diverse, quality, numeric data sets including microdata survey collection from DLI, demographic data from Statistics Canada and polling data from Gallup and other organizations.
- facilitates the exploration (searching, data manipulation, creation of summary statistics, graphing and export) of multiple, sophisticated data sets.
- Access is open to all OCUL institutions and is controlled by IP address
- <http://search1.odesi.ca/>

What is an RDC ?

- An RDC is a **secure Statistics Canada (Stat Can) research laboratory** physically located on a university campus to ensure that statistical microdata can be extensively analyzed by the research community across Canada without compromising the confidential nature of the data / information



RDC Microdata Resources / Links

Google: “McMaster RDC” for:

- Summary of data sets classified by type and status
 - Resources for selected surveys
 - Contact information and application procedures
-
- The RDC now has a vast data holding (e.g. 200+ surveys and datasets) on a wide rang of topics



A Brief List of Datasets

- Census master files (1911 to 2016)
- Canadian Community Health Survey (CCHS, annual)
- Canadian Health Measures Survey (CHMS)
- General Social Survey (GSS, annual)

- Canadian Survey on Disability 2012 (CSD)
- Canadian Tobacco Use Monitoring Survey: 1999-2012 (CTUMS)
- National Cannabis Survey (2018)
- Survey on Living with Neurological Conditions in Canada 2011 (SLNCC)

For full list, please visit [CRDCN's website](#)



Additional Datasets

Impressive growth in administrative and linked data being made available

- Canadian Cancer registry (CCR)
 - Vital Statistics Database: Births and Deaths
 - Ontario Child Health Survey (OCHS)
- ** Contact: Li Wang, RDC analyst**

Health linked datasets (linked by record ID to combine two or more datasets)

- Canadian births linked to the census
- Census linked to mortality, cancer and hospital data
- Canadian Cancer Registry linked to Vital Statistics Death Database
- CCHS linked to Hospital Discharge Data (DAD) and CCHS to Vital Stats Death



Additional Datasets

Coming soon:

- Longitudinal Immigration Database (IMDB) linked to hospital discharge (DAD)
- IMDB linked to CCHS
- Taxfiler Data (LAD) linked to hospital discharge (DAD)
- Others

**** Contact: Mustafa Ornek, RDC Analyst**



Access to Centre

Application Process:

- Academic, government funded or public sector access
- Link for “How to Apply”
<http://www.statcan.gc.ca/eng/rdc/process>
- Security Clearance and Orientation

McMaster RDC offers:

- Three analysts on-site for orientation and support
- 16 workstations with various statistical software
- Conference room



Contact Information

Data Liberation Initiative

- Mills Memorial Library,
Room L104/C
- 905 525-9140 Ext. 23848
- vivek@mcmaster.ca
- Hours of Service:
9:30 am – 1:00 pm
2:00 pm – 5:00 pm
- library.mcmaster.ca/data/

Research Data Centre

- L.R. Wilson Hall, Room
5806
- Analysts:
 - Peter Kitchen: rdc@mcmaster.ca
Ext. 27968
 - Mustafa Ornek: rdc2@mcmaster.ca
Ext. 27967
 - Li Wang: rdc4@mcmaster.ca
Ext. 27762
- Academic Director:
 - Michael Veall (rdc@mcmaster.ca)
- rdc.mcmaster.ca



Q & A / Discussion



Thank you

See you in the RDC ... SOON!



Telling Canada's
story in numbers