



Data Liberation Initiative (DLI) Resources for RDC Researchers

Vivek Jadon

Library Data Service, Mills Library

May 18, 2022

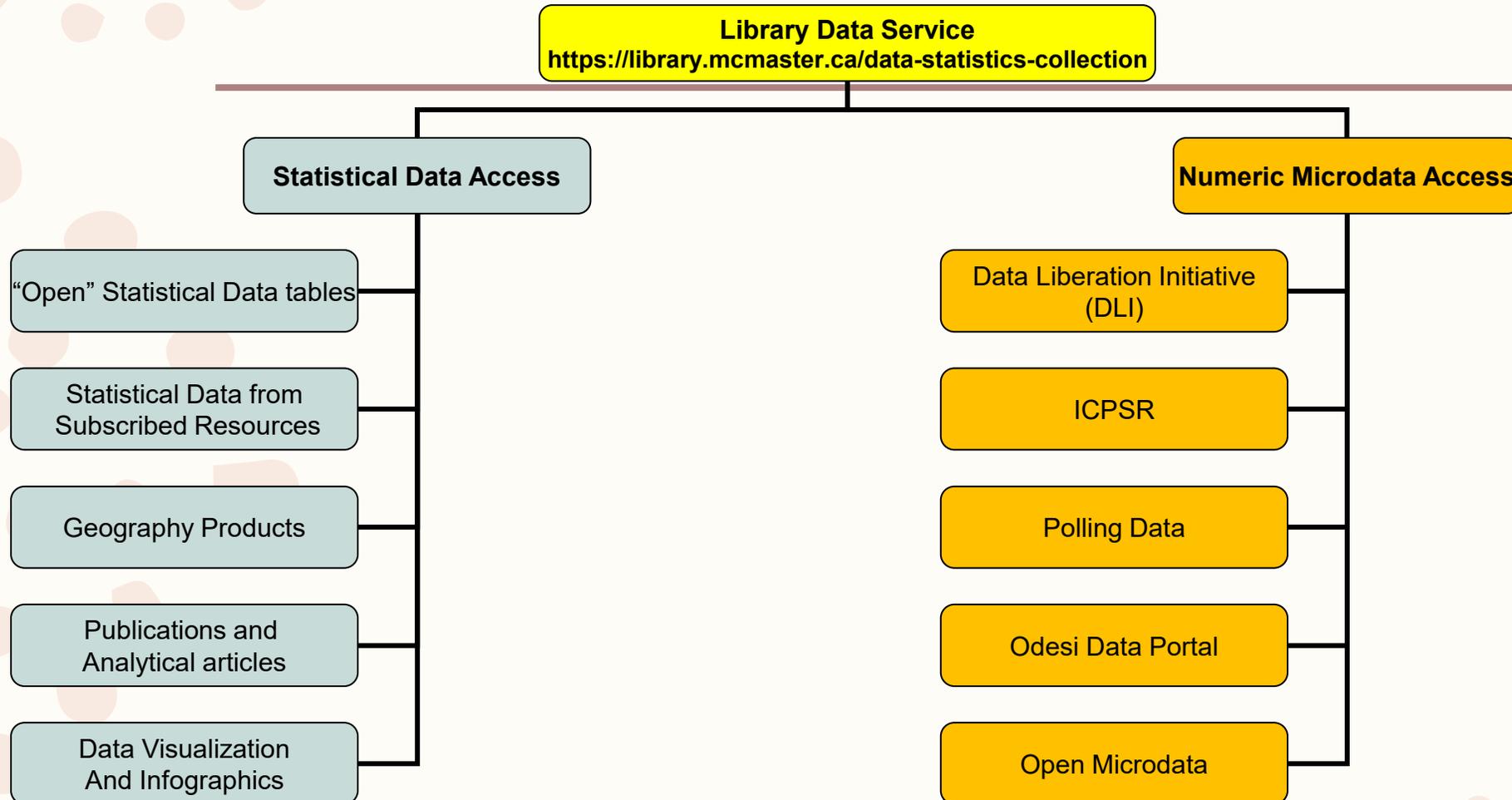
Session Outline

- Library Data Service - Remote Data Access
- Microdata Access Programs @ Statistics Canada
 - Data Liberation Initiative (DLI)
 - Real Time Remote Access (RTRA)
- Odesi Data Portal
- Data Analysis Support Hub (DASH) Program

Library Data Service

- The McMaster [Library Data Service \(LDS\)](#) promotes and facilitates the use of specialized numeric data products for **authorized McMaster users**.
- Library Data Service acquires data files as a member of two major consortia - Statistics Canada's Data Liberation Initiative (DLI) and Inter-University Consortium for Political and Social Research (ICPSR).
- The [data files](#) contain quantitative research data, including public use micro data, aggregate data and time-series databases.

Library Data Service “Remote” Data Access



CONTINUUM of MICRODATA ACCESS

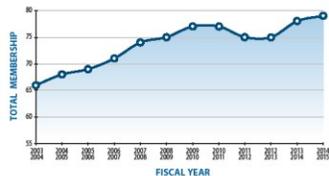


Nonrestricted Access

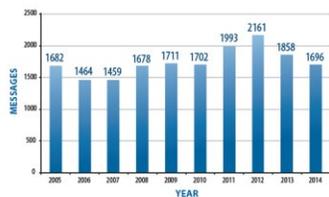
DATA LIBERATION INITIATIVE (DLI)

- 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

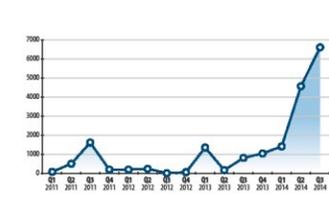


ACCESS IS A KEY PILLAR IN STATISTICS CANADA'S CORPORATE BUSINESS PLAN:

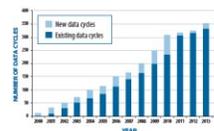
"Increasingly, modern policy research requires access not only to aggregated statistics but also to data at the level of the individual business, household or person. Statistics Canada is committed to removing obstacles to access while preserving the privacy and confidentiality of respondents."

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

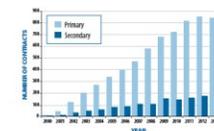
AVERAGE NUMBER OF SUCCESSFUL SUBMISSIONS



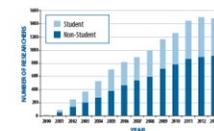
NUMBER OF DATA CYCLES IN THE RDCs



AVERAGE MONTHLY ACTIVE CONTRACTS



AVERAGE MONTHLY COUNTS OF RESEARCHERS



LOCATION OF RDCs ACROSS CANADA



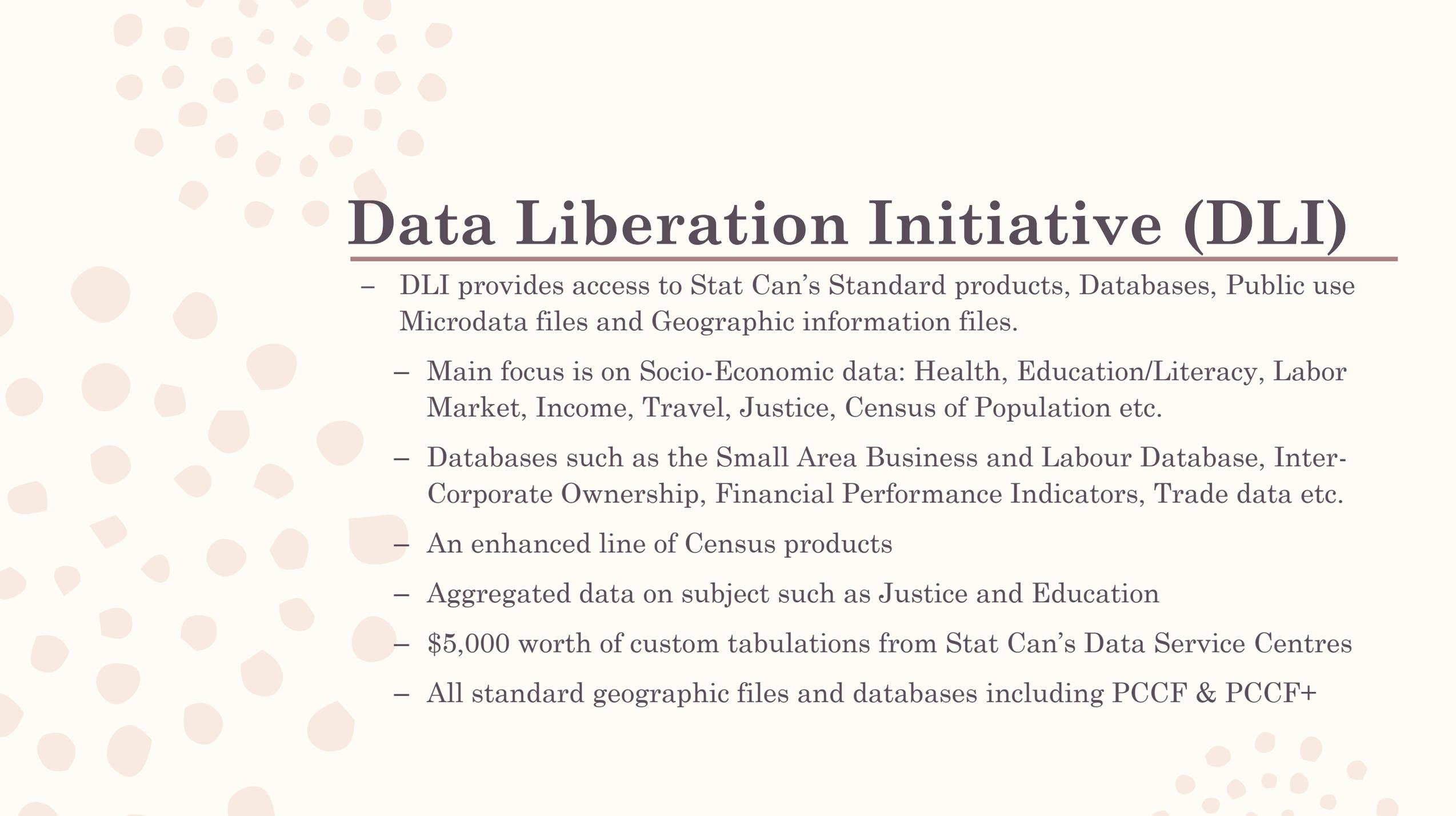
SECURITY REQUIREMENTS FOR AN RDC



FOR MORE INFORMATION:
www.statcan.gc.ca

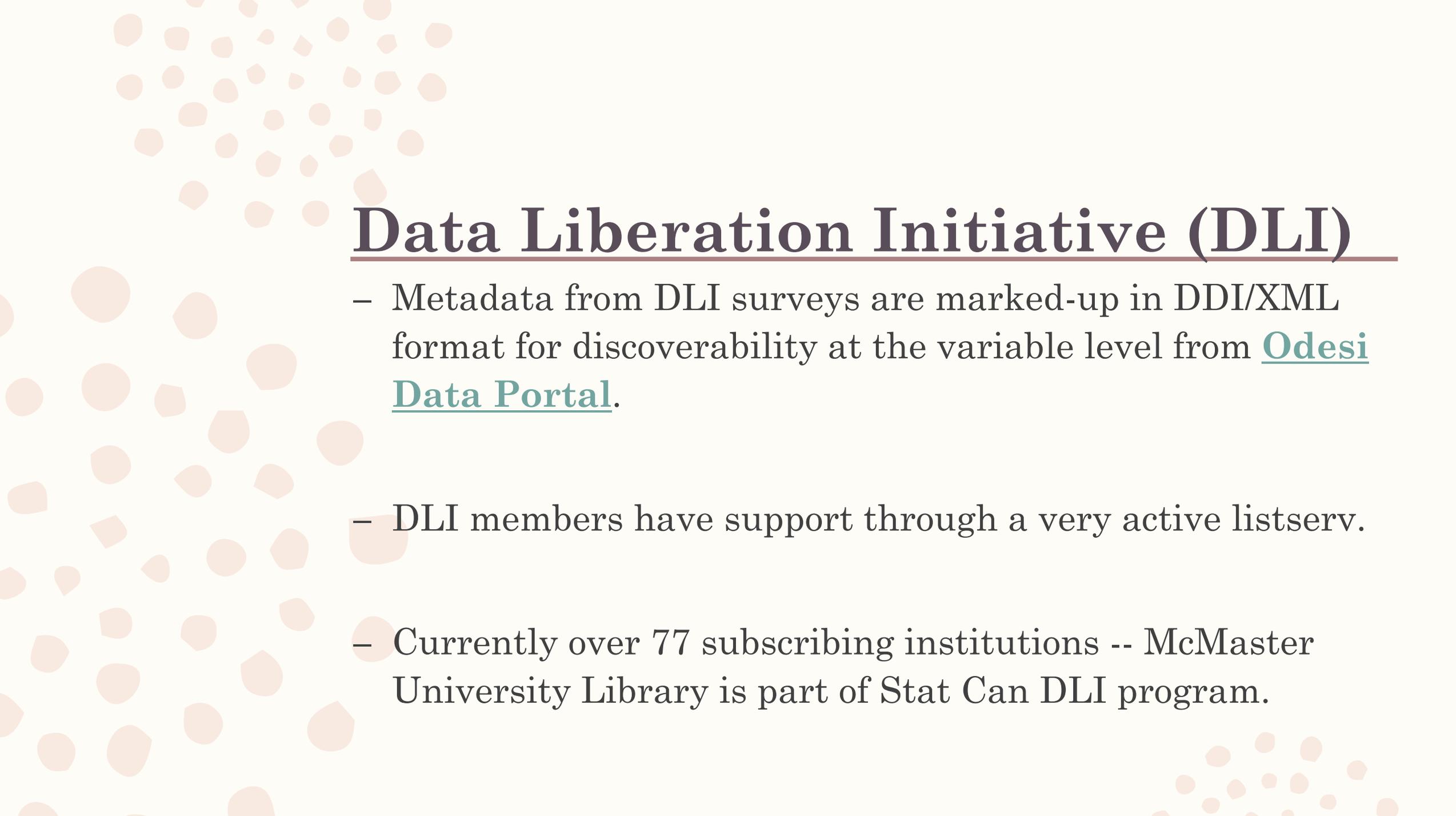
Microdata Access Programs

- “Public Use” Microdata access:
 - Data Liberation Initiative (DLI)
- “Restricted” Microdata access:
 - Real Time Remote Access (RTRA)



Data Liberation Initiative (DLI)

- DLI provides access to Stat Can's Standard products, Databases, Public use Microdata files and Geographic information files.
- Main focus 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
- \$5,000 worth of custom tabulations from Stat Can's Data Service Centres
- All standard geographic files and databases including PCCF & PCCF+



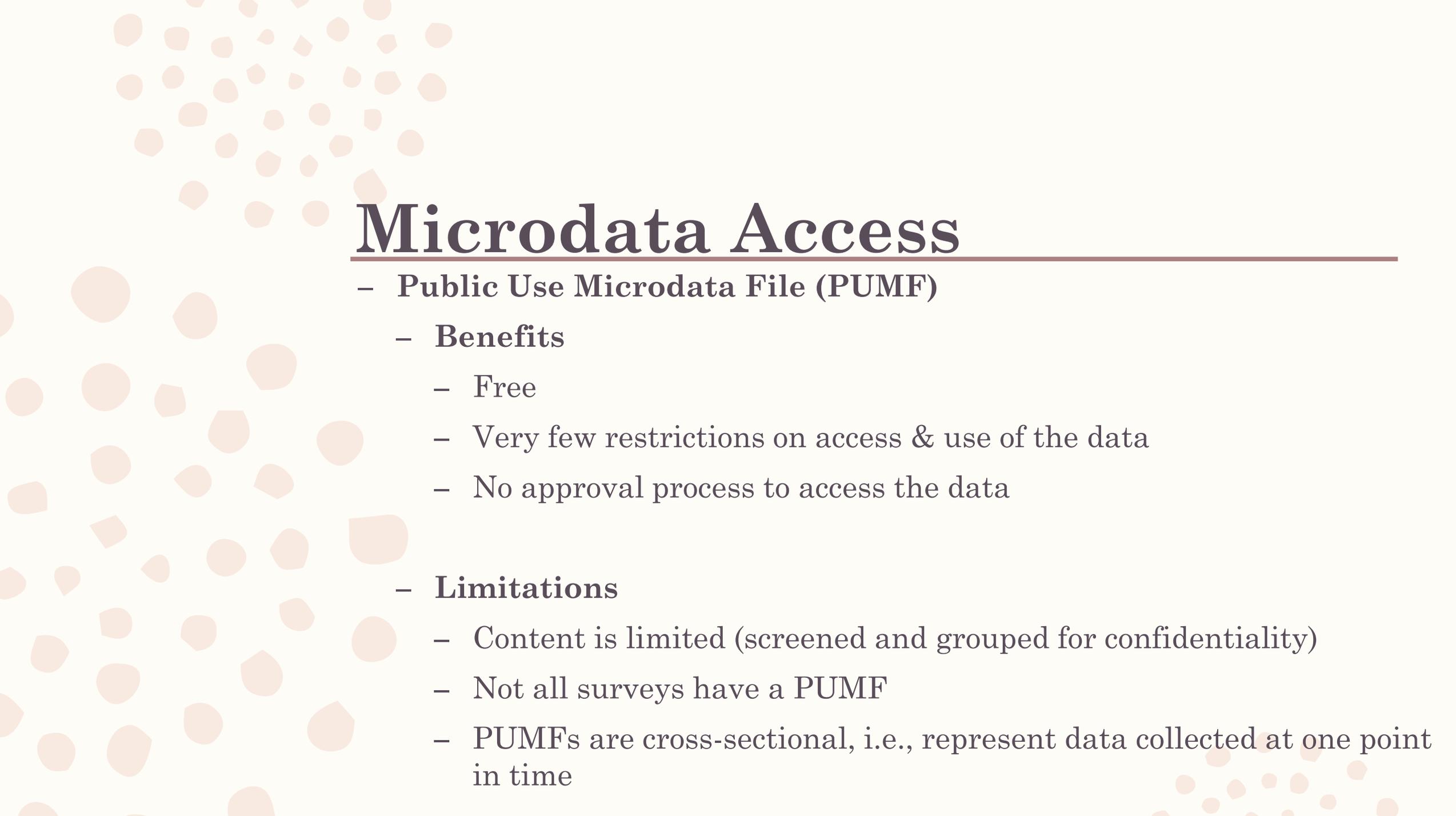
Data Liberation Initiative (DLI)

- Metadata from DLI surveys are marked-up in DDI/XML format for discoverability at the variable level from [Odesi Data Portal](#).
- DLI members have support through a very active listserv.
- Currently over 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



Most Popular DLI Collections

- Canadian Community Health Survey (CCHS)
- Labor Force Survey (LFS)
- Census of Population
- Discharge Abstract Database (DAD) -- CIHI
- General Social Survey (GSS)
- Survey of Household Spending (SHS)
- Canadian Tobacco, Alcohol and Drugs Survey (CTADS)
- National Household Survey (NHS)
- PCCF & PCCF+

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>

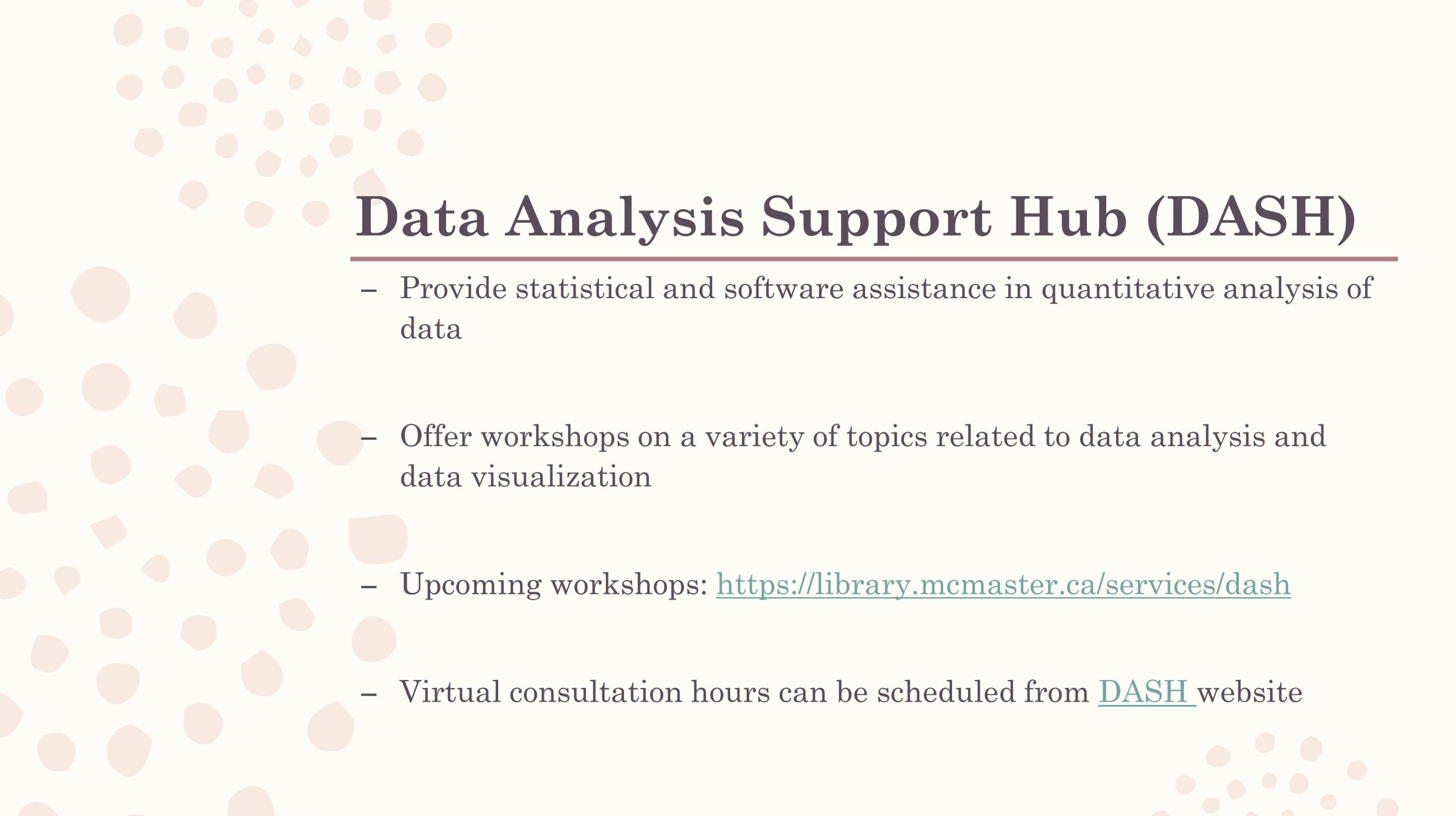
Inter-University Consortium for Social and Political Research (ICPSR)

- ICPSR: <http://www.icpsr.umich.edu/>
- ICPSR maintains and provides access to a vast archive of social science data for research and instruction.
- The ICPSR data contain mainly US survey statistics with some international content. The thematic categories lists data holdings on 17 broad subject areas such as education, health, justice etc.
- The ICPSR data holdings can be searched, downloaded and analyzed online. Users can either create a MyData account on ICPSR website or can login using Google/LinkedIn/ORCID accounts to download or analyze data online. Only authorized McMaster Users (current Faculty, Staff and Students) are allowed access to the data.
- ICPSR supplies data files for use with statistical software, such as SAS, SPSS, and Stata.

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://odesi.ca/>



Data Analysis Support Hub (DASH)

- Provide statistical and software assistance in quantitative analysis of data
- Offer workshops on a variety of topics related to data analysis and data visualization
- Upcoming workshops: <https://library.mcmaster.ca/services/dash>
- Virtual consultation hours can be scheduled from [DASH](#) website



Additional Resources

- Library Data Service Website
- Data & Statistics Collection
- Data and Statistics Libguides
- Data Analysis Support Hub (DASH)

Contact Information

Library Data Service

- Mills Memorial Library, Room L104/C
- (905) 525-9140 Ext. 23848
- vivek@mcmaster.ca

Hours of Service (2 days/week)

9:30 am – 1:00 pm

2:00 pm – 5:00 pm