

# Finding Canadian & International Microdata

<http://uoft.me/finding-microdata>

Lise Doucette, Data & Statistics Librarian

Nich Worby, Government Information and Statistics Librarian

[mdl@library.utoronto.ca](mailto:mdl@library.utoronto.ca)



UNIVERSITY OF TORONTO  
LIBRARIES

# Map & Data Library

## Services include

- Assistance with data discovery, use, and management
- Discovery tools to access numeric and geospatial data collections
- Specialized statistical and GIS software access
- Workshops and customized instruction sessions

## Get help

- In person: Mon-Fri, 11-5
- Software in computer lab: whenever Robarts is open
- By email: [mdl@library.utoronto.ca](mailto:mdl@library.utoronto.ca)



## Data

Datum (pl. data):

“factual information...used as a basis for reasoning, discussion or calculation”  
(Merriam-Webster)

Dataset:

A collection of facts (observations) that can be manipulated as a unit using a computer



## Microdata

A dataset containing the individual facts (observations) that were recorded.

For example:

- Individual responses to a survey
- Individual behavioral observations
- Hourly temperature readings at various weather stations
- Firm-level research & development spending
- Agriculture production by farm



# Microdata

Column: One variable  
or Question

Row: The unit of  
observation

Cell: One observation

\*cchs-82M0013-E-2013-2014-Annual-component\_F1.sav [DataSet1] - IBM SPSS Statistics Data Editor

	VERDAT E	ADM_RN O	GEOGRP V	GEOIDPM F	GEOIDBC HA	ADM_PR X	ADM_NO 9
1	20150427	1	Ontario	Durham	N/A	No	In Person
2	20150427	2	Quebec	Region de ...	N/A	No	NOT APPL...
3	20150427	3	Ontario	Peterborough	N/A	No	NOT APPL...
4	20150427	4	British Col...	Vancouver	Vancouver	No	Telephone
5	20150427	5	Ontario	City of Tor...	N/A	Yes	NOT APPL...
6	20150427	6	Ontario	Lambton	N/A	No	In Person
7	20150427	7	Quebec	Region de l...	N/A	No	Telephone
8	20150427	8	Ontario	Grey Bruce	N/A	No	In Person
9	20150427	9	Ontario	Hamilton	N/A	No	NOT APPL...
10	20150427	10	Ontario	Lambton	N/A	No	NOT APPL...
11	20150427	11	Nova Scotia	Cape Breton	N/A	No	NOT APPL...
12	20150427	12	Nova Scotia	Capital Dis...	N/A	No	In Person
13	20150427	13	Alberta	South Zone	N/A	No	NOT APPL...
14	20150427	14	Ontario	Ottawa	N/A	No	NOT APPL...
15	20150427	15	Manitoba	Northern R...	N/A	No	NOT APPL...
16	20150427	16	Ontario	Middlesex-...	N/A	No	NOT APPL...
17	20150427	17	Ontario	Ottawa	N/A	No	NOT APPL...
18	20150427	18	Quebec	Region de ...	N/A	No	NOT APPL...
19	20150427	19	New Bruns...	ZONE 3	N/A	No	NOT APPL...
20	20150427	20	British Col...	East Kooete...	Interior	No	In Person
21	20150427	21	Ontario	Peel	N/A	No	NOT APPL...
22	20150427	22	Quebec	Region de l...	N/A	No	NOT APPL...



## Data terminology

### Variable

- A characteristic being observed, that can take a range of values
- Observations may be numeric (e.g. temperature reading) or they may be descriptive (e.g. marital status)

### Case

- The unit of the analysis (e.g. if 20,000 people respond to a survey, then a microdata file will have 20,000 “cases” or rows)
- Often noted as  $n=20,000$ , where  $n$  is the number of cases



## Statistics

Collections of calculated figures produced through statistical inquiry – counts, tallies, totals, means/averages, percentages

For example:

- Percentage of teenagers who smoke cigarettes
- Median household income by city
- Approval rates for a political candidate



# Statistical table (aggregate data)

Job satisfaction by age group and sex, Canada, provinces and regions  
occasional (percent unless otherwise noted)

[Data table](#) [Add/Remove data](#) [Manipulate](#) [Download](#) [Related information](#) [Help](#)

The data below is a part of CANSIM table 121-0002. Use the [Add/Remove data](#) tab to customize your table.

Selected items [\[Add/Remove data\]](#)

Geography<sup>6</sup> = Canada

Age group<sup>8</sup> = Total, 15 years and over

Sex = Both sexes

Response <sup>7</sup>	Estimates <sup>9</sup>	2016
Total, job satisfaction	Number of persons (persons x 1,000)	15,167
	Percentage of persons	100.0
	Low 95% confidence interval, percent <sup>5</sup>	100.0
	High 95% confidence interval, percent <sup>5</sup>	100.0
Very satisfied or satisfied	Number of persons (persons x 1,000)	12,730
	Percentage of persons	83.9
	Low 95% confidence interval, percent <sup>5</sup>	82.0
	High 95% confidence interval, percent <sup>5</sup>	85.0

Total



Percentage





# Which do I need?

**Key factor: “How much/many?” vs “Why?”**

**“Does the Statistical Table I Need Exist”**

For example:

- “I need data on the number of Chinese immigrants in Toronto” ← Statistics
- “I’m exploring the relationship between income level and incidence of diabetes” ← Microdata



# Clarify your data needs

- What?
  - What concepts are you studying? What are some ways these could be operationalized into variables?
- Where?
  - Geographic areas/level of geographic detail
- When?
  - How current does the data need to be?
  - Do you need a time series?



# Develop a search strategy

- Make note of possible search terms for each variable
- Think about synonyms, broader & narrower terms
- Be prepared for trial and error, as every database works a bit differently



# Determine where to search

- What organization is most likely to produce what you're looking for?
- Where would their data likely be found? (data portal, organization's website?)
- Check the Map & Data Library website for guidance
- No luck? [Contact us!](#)



# Who collects microdata?

- Government bodies
- Researchers
- Commercial entities (e.g. polling firms)
- NGOs

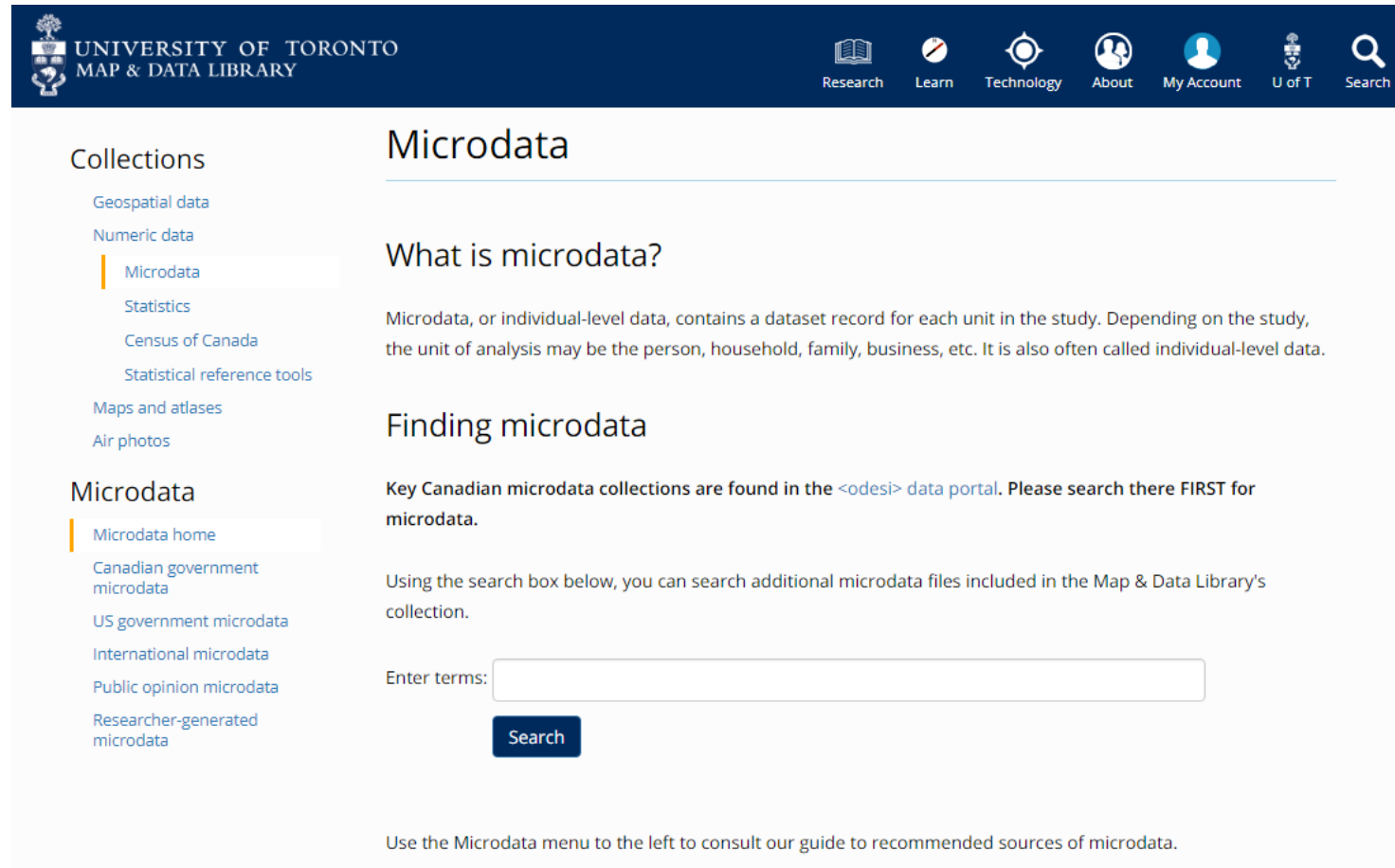
## Methods of data collection

- Enumeration (census)
  - An official count or survey every person/object
- Survey
  - A study which asks questions of a sample of the population
- Administrative
  - Counting things or people as they use government services or pass particular locations (e.g. hospital visits, imports/exports, tax returns filed)

Universities are secondary users – data is used for a different purpose than it was collected for



# MDL Website – Microdata Guide



The screenshot shows the MDL Website Microdata Guide page. At the top is a dark blue navigation bar with the University of Toronto Map & Data Library logo on the left and icons for Research, Learn, Technology, About, My Account, U of T, and Search on the right. Below the navigation bar is a left sidebar with a 'Collections' menu (Geospatial data, Numeric data, Microdata, Statistics, Census of Canada, Statistical reference tools, Maps and atlases, Air photos) and a 'Microdata' menu (Microdata home, Canadian government microdata, US government microdata, International microdata, Public opinion microdata, Researcher-generated microdata). The main content area has a 'Microdata' heading, a 'What is microdata?' section explaining that microdata is individual-level data, a 'Finding microdata' section with a search box and a 'Search' button, and a footer note to consult the guide for recommended sources.

**UNIVERSITY OF TORONTO**  
MAP & DATA LIBRARY

Research Learn Technology About My Account U of T Search

## Collections

- Geospatial data
- Numeric data
- Microdata**
- Statistics
- Census of Canada
- Statistical reference tools
- Maps and atlases
- Air photos

## Microdata

- Microdata home**
- Canadian government microdata
- US government microdata
- International microdata
- Public opinion microdata
- Researcher-generated microdata

## Microdata

### What is microdata?

Microdata, or individual-level data, contains a dataset record for each unit in the study. Depending on the study, the unit of analysis may be the person, household, family, business, etc. It is also often called individual-level data.

### Finding microdata

Key Canadian microdata collections are found in the [<odesi> data portal](#). Please search there **FIRST** for microdata.

Using the search box below, you can search additional microdata files included in the Map & Data Library's collection.

Enter terms:

**Search**

Use the Microdata menu to the left to consult our guide to recommended sources of microdata.

<https://mdl.library.utoronto.ca/collections/numeric-data/microdata>



# PUMFs

The screenshot shows the ODESI website interface. At the top, there is a navigation bar with the ODESI logo, a language selector for 'Français', and buttons for 'Browse', 'Contact', and 'My List 0'. Below the navigation bar is a search section titled 'Find data:'. It includes a search input field with the placeholder 'Enter Search Term', a location dropdown menu set to 'Anywhere', and a plus sign button. Below the search input is a 'Date:' section with two dropdown menus: 'Earliest' and 'Present'. Underneath is a 'Collections:' section with two columns of checkboxes. The first column, 'All of <odesi>', has four checked items: 'Statistics Canada (Microdata)', 'Statistics Canada (Aggregate data)', 'Public Opinion Polls', and 'Other Data'. The second column, 'Additional Collections:', has four unchecked items: 'Statistics Canada (Master File)', 'CORA', 'ICPSR', and 'Canadian Dataverses'. A blue 'Search' button is located at the bottom of the search section. To the right of the search section, there is a text box stating: 'ODESI contains over 3,535 datasets for the social sciences, with more than 13,000 additional dataset descriptions available for searching.'

## Learning Tools

[Video Tutorials](#)

[User Guide](#)

[FAQ](#)

## About

[About <odesi>](#)

[Who can Access it](#)

[Terms of Use](#)

## ODESI

- **Statistics Canada PUMFs**
- **Canadian public opinion data**





# DEMO: ODESI



# Exercise

A researcher is interested in investigating the social determinants of health. Specifically she would like data that allows her to investigate the correlation between socioeconomic status and diabetes diagnosis

- Using [ODESI](#), investigate what microdata we might have available on this topic. Find a dataset of interest and browse through the variables
- Bonus: can you create your own cross-tabulation showing diabetes by some socioeconomic variable (e.g. education level or income?)



# Statistics Canada Microdata

- Public Use Microdata Files (PUMFs) – Anonymized data files that U of T Students, Faculty and Staff can access
- Statistics Canada Master Files – Non-anonymized microdata with restricted access
- Synthetic Files – Fake data that allows for researchers to explore variables, but not conduct analysis



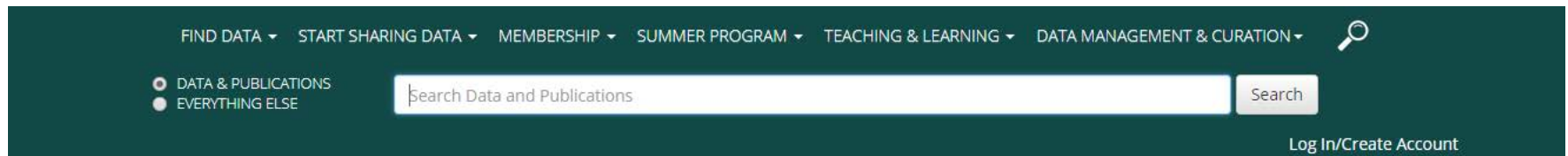
# Accessing Master Files

- Research Data Centre (RDC) – Need to submit application. You can only work with data in a RDC
- Real Time Remote Access (RTRA) – Need to submit application. Submit SAS code to create Statistical Tables
- Custom Tabulation – Pay Statistics Canada to create a Statistical Table for you



# US & international microdata

**ICPSR**: Archive containing US & International government & researcher data



The navigation bar is dark green with white text. It features a series of menu items: FIND DATA ▾, START SHARING DATA ▾, MEMBERSHIP ▾, SUMMER PROGRAM ▾, TEACHING & LEARNING ▾, and DATA MANAGEMENT & CURATION ▾. A magnifying glass icon is on the right. Below the menu items, there are two radio buttons: 'DATA & PUBLICATIONS' (selected) and 'EVERYTHING ELSE'. A search bar with the placeholder text 'Search Data and Publications' and a 'Search' button is positioned to the right of the radio buttons. In the bottom right corner, there is a link for 'Log In/Create Account'.



UNIVERSITY OF TORONTO  
LIBRARIES

# Exercise

Using [ICPSR](#), find a US study, published in the last 10 years, that contains variables on diabetes diagnosis and socioeconomic status.

- Bonus: if your study has online analysis available, create your own cross-tabulation showing diabetes by some socioeconomic variable



# US & international microdata

**IPUMS**: Harmonized international census data and US government microdata

## IPUMS

[ABOUT](#) | [SUPPORT](#) | [PROJECTS](#) | [BIBLIOGRAPHY](#) | [BLOG](#)

IPUMS provides census and survey data from around the world integrated across time and space. IPUMS integration and documentation makes it easy to study change, conduct comparative research, merge information across data types, and analyze individuals within family and community context. Data and services available free of charge.



U.S. Census and American Community Survey microdata from 1850 to the present.

[VISIT SITE](#)



Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present.

[VISIT SITE](#)



Census microdata covering 82 countries from 1960 to the present. [IPUMS NAPP](#) offers microdata from the 19th and early 20th centuries.

[VISIT SITE](#)



Health survey data for Africa and Asia, including harmonized data collections for [DHS](#) and [PMA2020](#).



Tabular U.S. Census data and GIS boundary files from 1790 to the present.



Integrated data on population and the environment from 1960 to the present.

### RESEARCH AWARDS

2017 IPUMS Research Award submissions are now open.

[LEARN MORE](#)

### SUPPORT IPUMS

Your financial support helps us bring research data to you.

[DONATE TODAY](#)

### CALENDAR

Allied Social Sciences Association



UNIVERSITY OF TORONTO  
LIBRARIES

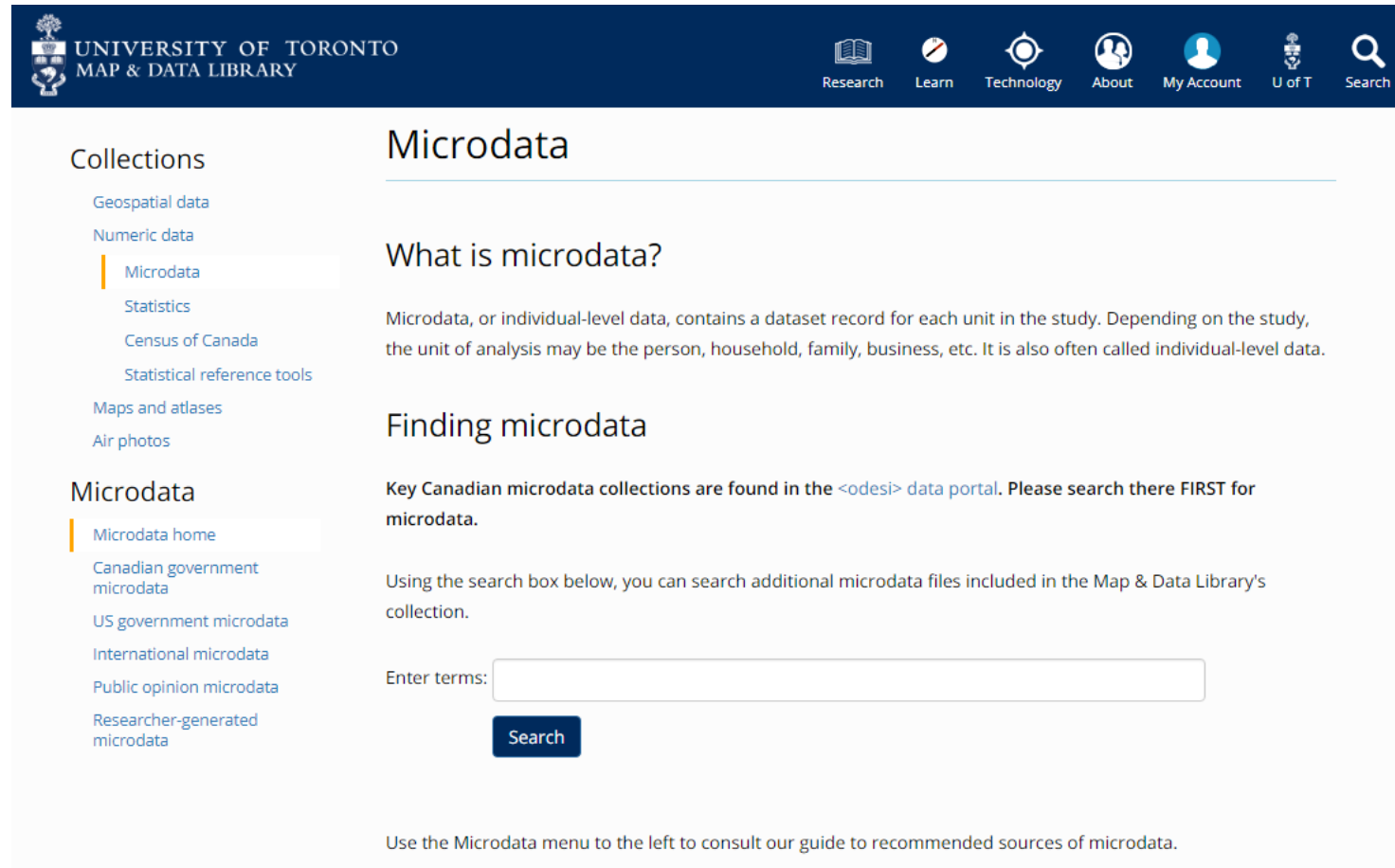
# Exercise

Select an IPUMS tool of interest to you and explore the available variables. Create a data extract and/or play with the “analyze data online” tools.





# Find more sources – Microdata Guide



The screenshot shows the University of Toronto Map & Data Library website. The top navigation bar includes the library logo and icons for Research, Learn, Technology, About, My Account, U of T, and Search. The main content area is titled "Microdata" and contains a "What is microdata?" section, a "Finding microdata" section with a search box, and a "Collections" sidebar on the left. The sidebar lists various data categories, with "Microdata" highlighted. The search box is labeled "Enter terms:" and has a "Search" button below it.

**UNIVERSITY OF TORONTO**  
MAP & DATA LIBRARY

Research Learn Technology About My Account U of T Search

## Collections

- Geospatial data
- Numeric data
- Microdata**
- Statistics
- Census of Canada
- Statistical reference tools
- Maps and atlases
- Air photos

## Microdata

- Microdata home**
- Canadian government microdata
- US government microdata
- International microdata
- Public opinion microdata
- Researcher-generated microdata

## Microdata

### What is microdata?

Microdata, or individual-level data, contains a dataset record for each unit in the study. Depending on the study, the unit of analysis may be the person, household, family, business, etc. It is also often called individual-level data.

### Finding microdata

Key Canadian microdata collections are found in the [<odesi> data portal](#). Please search there **FIRST** for microdata.

Using the search box below, you can search additional microdata files included in the Map & Data Library's collection.

Enter terms:

**Search**

Use the Microdata menu to the left to consult our guide to recommended sources of microdata.

<https://mdl.library.utoronto.ca/collections/numeric-data/microdata>



# Questions?



UNIVERSITY OF TORONTO  
LIBRARIES

# Evaluation / Post-it note

1. One useful thing you learned in this workshop
2. One recommendation for us for the next time we offer this workshop

