

# Open Data for Epidemiological Modelling

Noah Bolohan, Analyst  
[noah.bolohan@statcan.gc.ca](mailto:noah.bolohan@statcan.gc.ca)



Delivering insight through data for a better Canada



Statistics  
Canada

Statistique  
Canada

Canada

# Summary

- Introduce use case: ODFEM
- Open data - overview
- Specifics: Census / CODR / LODDE
- Open data applied to ODFEM



# Use-case: ODFEM

- Open Data Framework for Epidemiological Modelling

- Inspired by June (UK)

**ROYAL SOCIETY  
OPEN SCIENCE**

Open Access  
Check for updates  
**View PDF**

Tools Share  
Cite this article

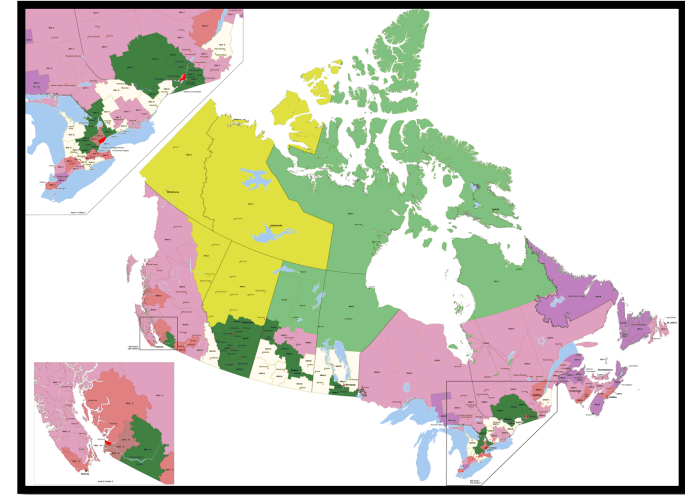
Research articles  
**June: open-source individual-based epidemiology simulation**  
Joseph Aylett-Bullock<sup>†</sup>, Carolina Cuesta-Lazaro<sup>†</sup>, Arnau Quera-Bofarull<sup>†</sup>, Miguel Icaza-Lizaola<sup>‡</sup>, Aidan Sedgewick<sup>‡</sup>, Henry Truong<sup>‡</sup>, Aoife Curran, Edward Elliott, Tristan Caulfield, Kevin Fong, Ian Vernon, Julian Williams, Richard Bower and Frank Krauss<sup>✉</sup> See fewer authors ^

Published: 07 July 2021 | <https://doi.org/10.1098/rsos.210506>



## Use-case: ODFEM

- Geographically granular national epidemiological model
- Disease transmission from interpersonal contact
- Curated collection of Canadian open data



# Data overview

## Census Public Use Microdata



## Linkable Open Data Environment



## CODR Tables

Work status	Total, work status				
Age group	Total, age group				
Sex	Total, sex				
Statistics	Number of educators <sup>6, 7</sup>				
Geography	2018 / 2019	2019 / 2020	2020 / 2021	2021 / 2022	2022 / 2023
	Number				
Canada ( <a href="#">map</a> )	401,286	402,666	409,785	410,142	413,667
Newfoundland and Labrador <sup>8</sup> ( <a href="#">map</a> )	5,367	5,343	5,595	5,490	5,502
Prince Edward Island <sup>8</sup>	1,695	1,701	1,710	1,537	1,700

- [CANFED](#)
- [Boundary Files](#)
- [General Social Survey](#)
- [And more...](#)



# Data overview – Census PUMF

**Census of Population: Public Use Microdata Files**

Release date: September 12, 2023 | Updated on: March 20, 2024

Census Public Use Microdata Files (PUMFs) provide access to non-aggregated data covering the population. It is a comprehensive social, demographic and economic database about Canada, showing a wide range of characteristics on the population. These files enable the study of individuals, families, and communities. Geographic identifiers have been restricted to ensure respondents' anonymity. This content is a useful tool for policy analysts, pollsters, social researchers and anyone interested in modeling and regression analysis using Census microdata.

These products contain the data file, user documentation and license agreements. It is important to note that these products require knowledge of data manipulation packages (or software) such as SAS, SPSS or Stata.

**2021**

- [Individuals file, 2021 Census of Population](#)

**2016**

- [Individuals file, 2016 Census of Population](#)
- [Hierarchical file, 2016 Census of Population](#)

1. Census PUMF landing page

## 2. .zip contents

- English
- Français
- data\_donnees\_2021\_ind.dat
- data\_donnees\_2021\_ind\_v2.csv

- Questionnaires
- Reference Guides
- 2021 Census Individuals PUMF User Guid...

## 3. 'English' contents

## 4. 'Reference Guides' contents

- cen\_2021\_age\_sex\_gender.pdf
- cen\_2021\_commuting.pdf
- cen\_2021\_dwelling.pdf
- cen\_2021\_education.pdf
- cen\_2021\_Ethnic\_Cultural\_origin.pdf
- cen\_2021\_families.pdf
- cen\_2021\_housing.pdf
- cen\_2021\_immigration.pdf
- cen\_2021\_Income.pdf
- cen\_2021\_Income-eng.pdf
- cen\_2021\_indigenous.pdf
- cen\_2021\_Labour.pdf
- cen\_2021\_languages.pdf
- cen\_2021\_military.pdf
- cen\_2021\_Minority Official Language.pdf
- cen\_2021\_mobility\_migration.pdf
- cen\_2021\_religion.pdf
- cen\_2021\_visible Minority and Populatio...



# Data overview – CODR Tables

**Data**

Filter results by ⓘ

Sort by: date  ⓘ Show: 10 entries

Keyword(s)

Search title, description, etc.

**Census of Population**

Census (2968)

**Subject**

Agriculture and food (710)

Business and consumer services and culture (428)

Business performance and ownership (1240)

Children and youth (583)

Construction (375)

[More](#)

**Geography**

Canada (6,035)

Search by ve ⓘ

Statistics Canada's Trust Centre

[How to access microdata](#)

[Customized products and services](#)

All (12,152) Tables (11,421) Profiles of a community or region (284) Thematic maps (71) Public microdata

1. [PASSAGES microsimulation model](#)

Table: 89-26-0006

Description: PASSAGES is an open-source dynamic microsimulation model aimed at supporting policy analysis and research relating to Canadian retirement income system outcomes at the individual and...

[More](#)

Release date: 2024-10-18

## Labour force characteristics by province, monthly, seasonally adjusted<sup>1, 2, 3, 4</sup>

Frequency: Monthly

Table: 14-10-0287-03

Release date: 2024-10-11

Geography: Canada, Province or territory

▼ Customize table

Sex :  Age group :  Reference period

Time :

ⓘ

Didn't find what you're looking for? [View related tables, including other calculations and frequencies](#)

Showing 99 records

Sex	Both sexes							
Age group	15 years and over							
Statistics	Estimate							
Data type	Seasonally adjusted							
	September 2023 to September	September	September	September	September	August 2024 to September	August 2024 to September	September 2023 to September
Labour force								





# Data overview - LODÉ

## The Linkable Open Data Environment

### Overview

The Linkable Open Data Environment (LODE) is an exploratory initiative that aims at enhancing the use and harmonization of open micro data primarily from municipal, provincial and federal sources.

The results are a collection of datasets released under a single open data license ([Open Government Licence - Canada](#)), as well as open source tools used to process the data, and collaborations in an open space.

The LODÉ is implemented by the Data Exploration and Integration Lab (DEIL). Thus, it is an ongoing and exploratory initiative, developed in an open space, as an open project.



## Open databases released

### The Open Database of Businesses



[The Open Database of Businesses](#) including names and characteristics of businesses.

Release date:

- November 28, 2023

### The Open Database of Infrastructure



[The Open Database of Infrastructure](#) including tunnels, solid waste, water, stormwater, and infrastructure providers.

Release date:

- May 8, 2023

### The Open Database of Greenhouses



[The Open Database of Greenhouses](#)





# Application to ODFEM

## Create the world



### Input data:

- Academia
- CANFED
- CIHI
- LODE
- StatCan

### World prep:

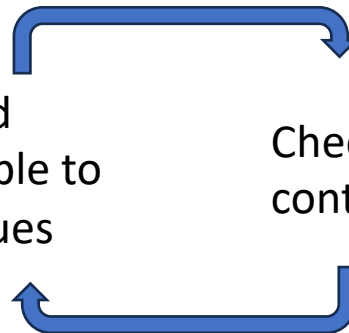
- Venues
- Population
- Households

## Run the simulation

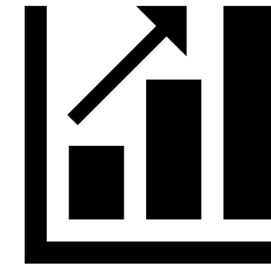


Send  
people to  
venues

Check  
contacts



## Output results



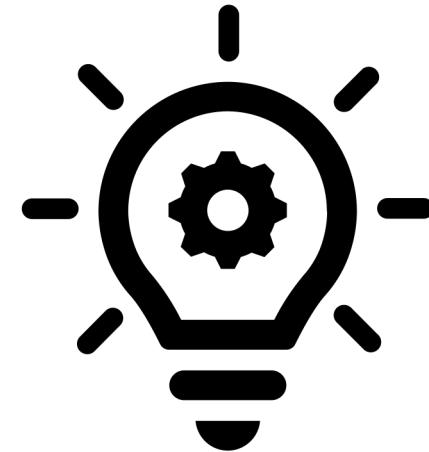
- World summary
- All contacts
- Model-generated contact matrices

Data source	Use in ODFEM
<a href="#">2021 Census PUMF</a>	<ul style="list-style-type: none"> <li>• Demographic profile</li> <li>• Household structure</li> </ul>
<a href="#">CODR Tables</a>	<ul style="list-style-type: none"> <li>• <a href="#">Company Sectors / Size</a></li> <li>• <a href="#">Educators</a></li> <li>• <a href="#">Labour Force Characteristics</a></li> </ul>
<a href="#">LODE</a>	<ul style="list-style-type: none"> <li>• <a href="#">Educational Facilities</a></li> <li>• <a href="#">Healthcare Facilities</a></li> <li>• <a href="#">Recreational and Sports Facilities</a></li> </ul>
<a href="#">CIMD</a>	<ul style="list-style-type: none"> <li>• Population per dissemination area</li> <li>• Socio-economic index</li> </ul>
<a href="#">General Social Survey</a>	<a href="#">Activity Frequency</a>
<a href="#">Boundary Files</a>	<ul style="list-style-type: none"> <li>• <a href="#">Geography correspondence files</a></li> <li>• <a href="#">Health Regions</a></li> </ul>
<a href="#">CANFED</a>	Grocery stores and restaurants



# Summary of ODFEM

- Open-source (!!!)
- Scope: national to public health unit\*
- Customizability/adaptability
- Generic outputs = wider usage
- All done using open data



\* When possible. Not always available, e.g., data limitations, changing structure

# Thank you for your time!



Please send questions to myself at: [noah.bolohan@statcan.gc.ca](mailto:noah.bolohan@statcan.gc.ca)

