

# Statistics Canada - Linkable File Environments, an Overview

(November 2023)



Statistics  
Canada

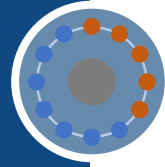
Statistique  
Canada

Canada





# Presentation Outline



**Data Linkable Environments at Statistics Canada: B-LFE; SDLE and LODE**

---



**How to fill in the Data Gaps Using Statistics Canada Linkable data Environments**

---



**Data Linkage Access Process**

---



# Linkable File Environments at Statistics Canada



# What's a Linkable File Environment

- **Definition:**

The Data Linkable Environment provides information about business, individuals and economic issues by facilitating the creation of datasets combining survey and administrative data.

Data linking takes personal applicant information from one dataset and connects it to other datasets about the matching individuals or businesses to enable anonymized statistical intersectional analysis, comparisons, and evaluation.

- **Advantages:**

Respond to data Gaps; Low cost; Low respond burden; Coverage and frame continuously updated

- **Legal Framework:**

Mandate; approvals; administrative data and confidentiality

- **Objective:**

Respond to policymakers and researcher to their analytical needs using micro-data



# Linkable File Environments at Statistics Canada

- **Business Linkable File Environment(B-LFE)**

*Objective:* Provide micro data demographic and characteristics information on businesses, Entrepreneurship and Ownership **at the enterprise level.**

*Link web page:* <https://www.statcan.gc.ca/en/about/statcan/lfe>

- **Social Data Linkage Environment (SDLE)**

*Objective :* Enable linkage at the **individual level** across multiple data sets in the social domain to facilitate pan-Canadian social and economic statistical research.

*Link web page:* <https://www.statcan.gc.ca/en/sdle/overview>

- **Linkable Open Data Environment (LODE)**

*Objective :* Provide open micro data to the public on a variety of services, amenities and infrastructures at the **location and geographic level.**

*Link web page:* <https://www.statcan.gc.ca/en/lode>



## Advantages of Linked Data

- **Fills data gaps**
- **Reduces response burden**
- **Reduces record linkage costs**
- **Continuously updated frame, with new data and sources**
- **Allows coverage of the entire business and individual population over long periods**



# How to fill in the Data Gaps Using Statistics Canada Linkable Environments



# There is a growing demand for information requiring simultaneously business, individual and geographic characteristics

## □ How to reconcile information from multiple data sources:

- Require a common identifier (Key)
- Require a consistent data environment (A common metric -Enterprise; Individual; geographic location...)

## □ Development of new indicators:

- Ownership characteristics; quality of life; equity; diversity; inclusion indicators; New taxonomies and more...







# The Business Linkable File Environment (B-LFE)



# Business Linkable File Environment (B-LFE)

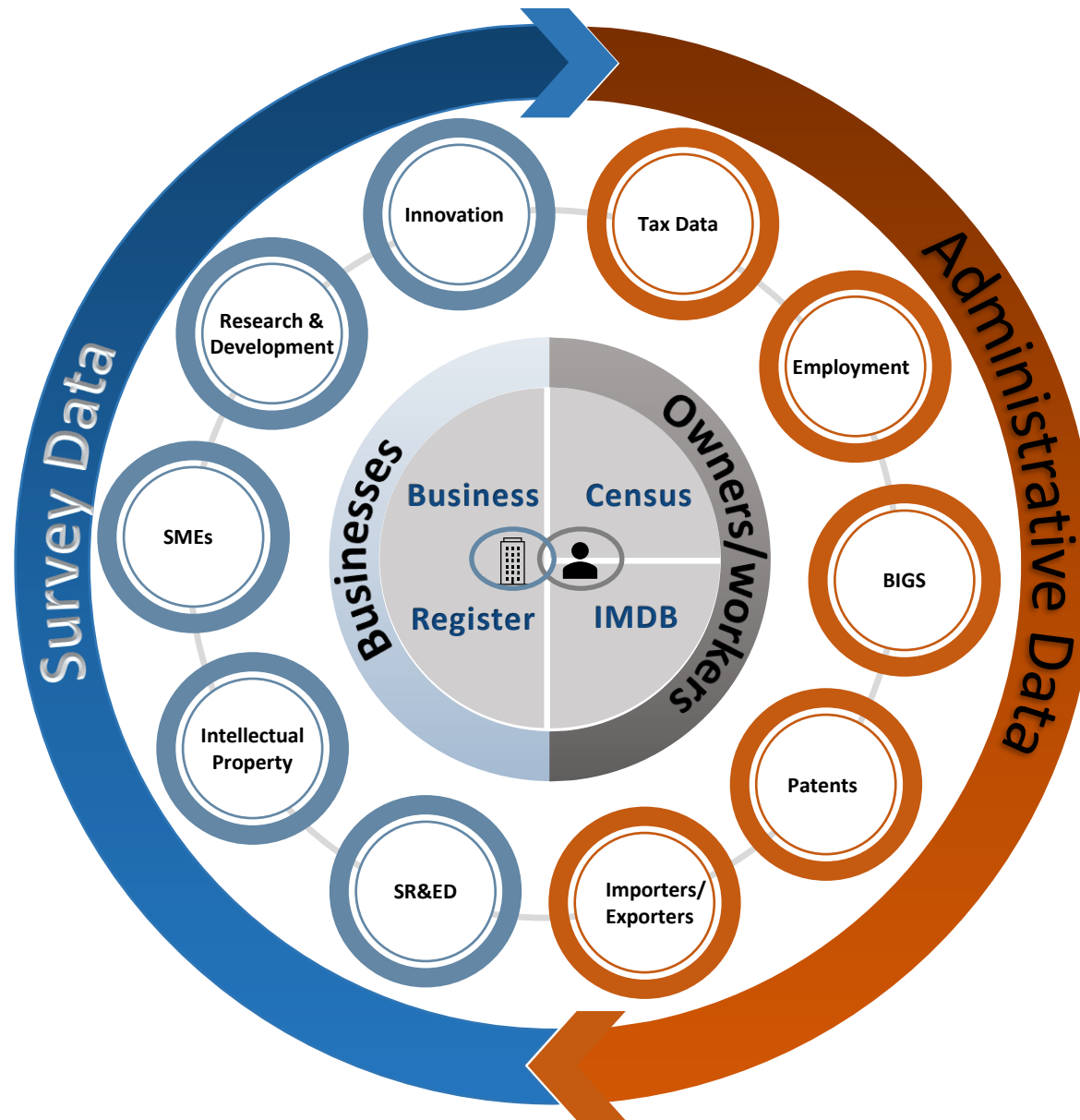
Covers all Canadian businesses

Over 30 data sources

Facilitates the creation of linked survey and administrative data files to inform on business and economic issues

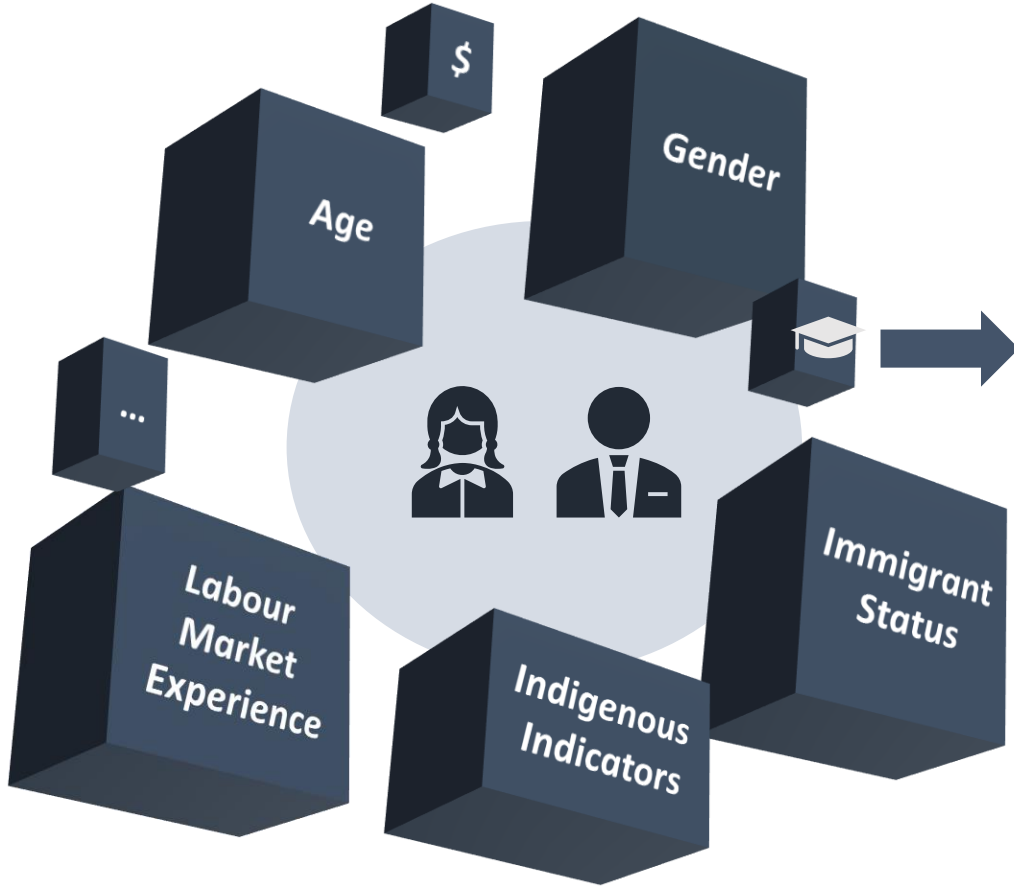
Analytical Opportunities:

- Assess business performance for reporting purposes
- Program impact evaluation
- Customized tabulations
- Gender and Diversity Indicators



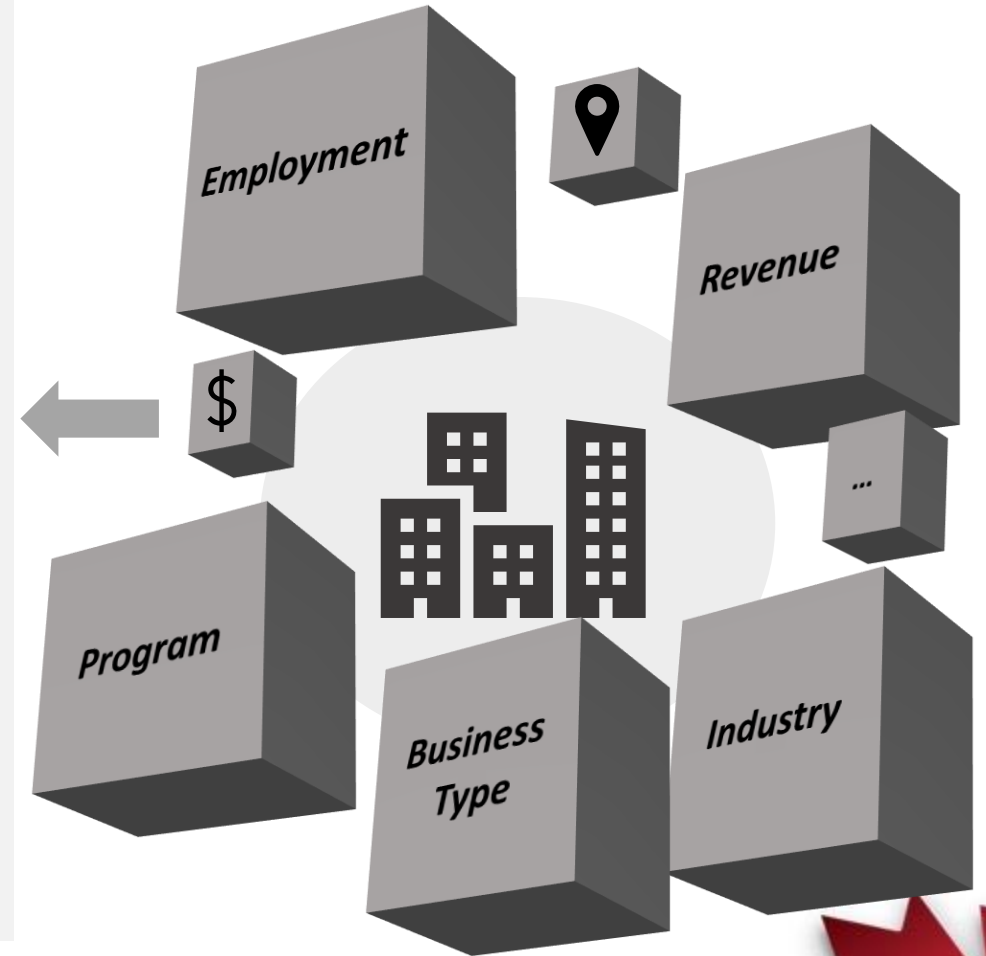
# What data is available for analytical insights and decision making?

## Individual Characteristics

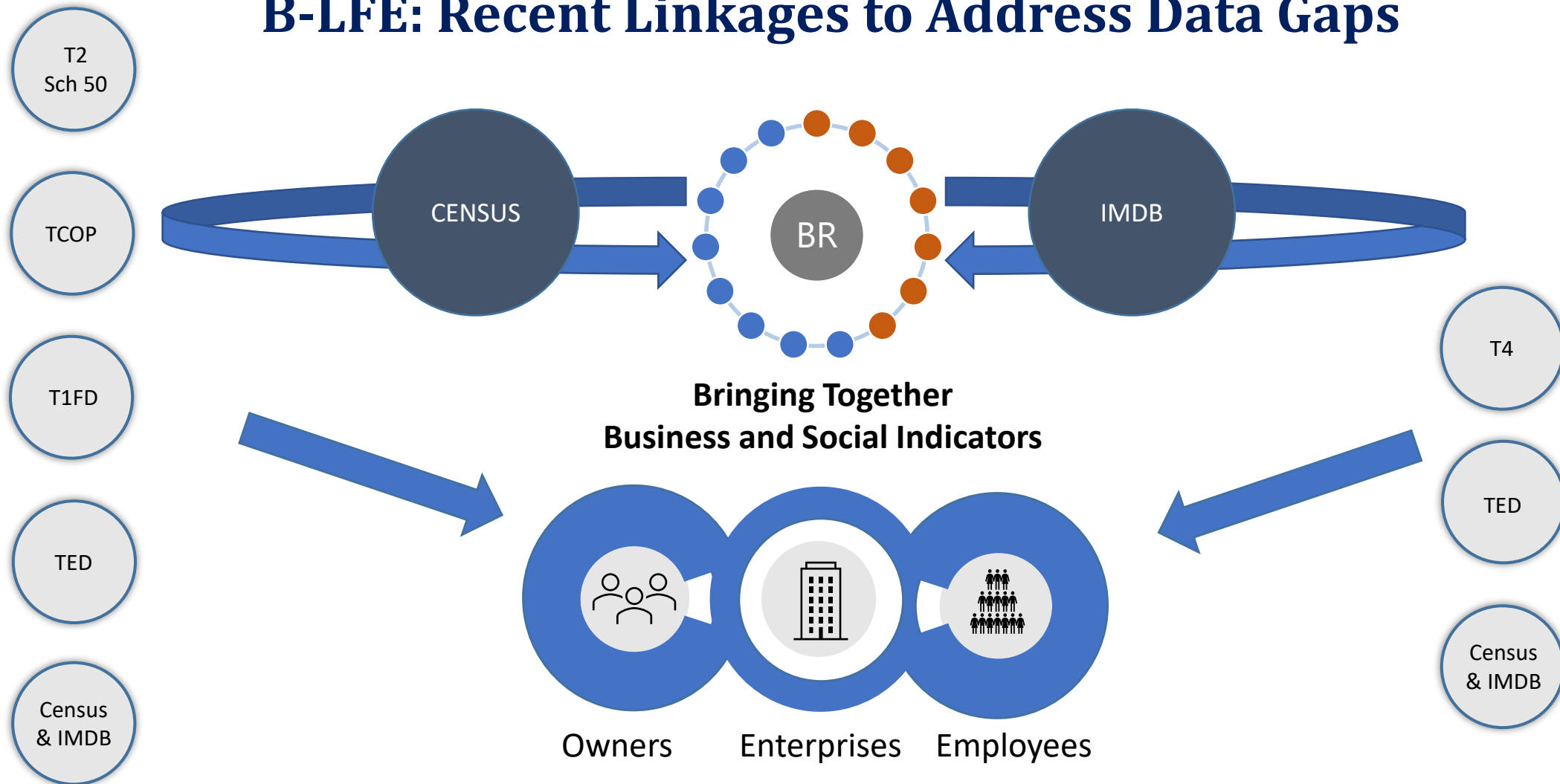


Understanding Not ONLY the impact of programs on businesses but also allows for analysis of employer-employees and ownership individual characteristics

## Business Characteristics



# B-LFE: Recent Linkages to Address Data Gaps



## Sch 50

List of shareholders who hold 10% or more of all private corporations' shares

## TCOP

Files of BNs and contact details of owners (includes unincorporated businesses)

## T1FD

Owners of unincorporated businesses filing T1 income

## TED

Administrative data for individuals (tax filers)

## T4

Administrative data on salary, wages, and taxable benefits paid to employees





## BIGS Products – using the B-LFE for Gender and Diversity Analysis

### Other upcoming releases:

- Analysis of Black-owned Businesses Receiving Federal Innovation and Growth Support (BIGS), 2020
- A Portrait of Official-Language Minority (OLM) Owned Businesses Receiving Federal Innovation and Growth Support (BIGS), 2020

### FEDERAL BUSINESS INNOVATION AND GROWTH SUPPORT

#### TO OFFICIAL LANGUAGE MINORITY OWNED BUSINESSES, 2020



In 2020, **344,225** businesses were OLM-owned<sup>1</sup>, representing **6.3%** of all businesses<sup>2,3</sup> in Canada. Of these businesses, **1,775** (0.5%) received **\$437,595,990** (12.7%) of all business innovation and growth support.

Compared with unsupported OLM-owned businesses, **supported<sup>4</sup> OLM-owned businesses** were:

**10.1 times**

more likely to be  
**mid-sized businesses**  
(100 to 499 employees)

**33.1 times**

more likely to be  
**large businesses**  
(500 or more employees)

**4.8 times**

more likely to report  
making at least  
**\$2 million in revenue**

**1.9 times**

less likely to be  
owned by women

Nearly half (**48.7%**) of all supported OLM-owned businesses were concentrated in professional, scientific and technical services or manufacturing, compared with **14.9%** of unsupported OLM-owned businesses.

Compared with OLM-owned businesses that did not receive support, **supported OLM-owned businesses** were:



**10.5 times**  
more concentrated in  
**manufacturing**



**8.6 times**  
more concentrated in  
**mining, quarrying, and  
oil and gas extraction**



**3.2 times**  
more concentrated in  
**information and  
cultural industries**



**22.6 times**  
less concentrated in  
**real estate, and  
rental and leasing**

#### Notes:

- In this infographic, official language minority (OLM) owned businesses refer to businesses whose owner has English as their first official language spoken (FOLS) in Quebec and French as their FOLS in the rest of Canada. Those with both English and French as their first official languages spoken were equally redistributed among the English-only and French-only categories.
- Business owners are primary owners of new, alive, or alive, but inactive businesses operating in Canada, including all those with self-employment income, that are linked to a 2021 Census record.
- The primary business owner is the individual most likely to be the business owner (or manager) among all known owners and office bearers. If multiple owners exist, the larger share of business income determines the primary owner.
- The primary owner of a supported business is the ultimate recipient of federal business innovation and growth support.

Sources: Statistics Canada, Business Linkable File Environment, 2020, Business Innovation and Growth Support, 2020, and Census of Population, 2021.

© His Majesty the King in Right of Canada, as represented by the Minister of Industry, 2023

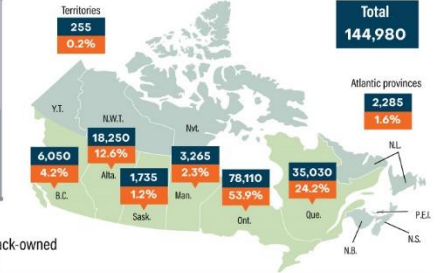


## BLACK-OWNED BUSINESSES in Canada, 2020

→ Distribution of Black-owned<sup>1</sup> businesses, by region<sup>2</sup>

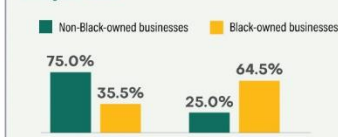
In 2020, there were an estimated **144,980** Black-owned businesses in Canada, representing **2.4%** of the total number of businesses in the country.<sup>3</sup>

Black-owned businesses are more likely to be owned by immigrants, younger people, and the self-employed.

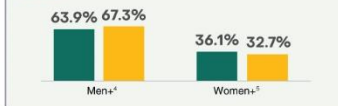


→ Percentage of Black-owned and non-Black-owned businesses, by owner characteristics

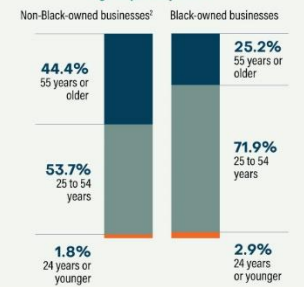
#### Immigrant status



#### Gender



#### Age of primary owner



→ Percentage of Black-owned and non-Black-owned businesses, by business characteristics

#### Self-employed (sole proprietorship)

**59.6%**  
NON-BLACK-OWNED  
BUSINESSES

**69.5%**  
BLACK-OWNED  
BUSINESSES



In 2020, close to 1% of federal business innovation and growth support (BIGS) went to Black-owned businesses. This represents approximately \$42 million of the total \$3.8 billion in federal funding through BIGS.

Black-owned businesses that received federal support through BIGS were more likely to be owned by immigrants (56.7%) than non-Black-owned businesses that received support (22.0%).

#### Notes:

- Business owners are defined as the primary owners of active businesses operating in Canada, including those with self-employment income. The primary business owner is the individual most likely to be the business owner (or manager) among all known owners and office bearers. This primary owner is a single person with the largest ownership share. If multiple owners meet this criterion, the primary owner is the person with the largest total income.
- Percentages do not always add up to 100% due to rounding.
- Owners who could not be linked to a census record were excluded.
- This category includes men and may also include some non-binary persons.
- This category includes women and may also include some non-binary persons.

Sources: Statistics Canada, Business Linkable File Environment, 2020, Business Innovation and Growth Support, 2020, Census of Population, 2021, 2006, 2010, 2011, National Household Survey, 2011, and Longitudinal Immigration Database, 2020.

© His Majesty the King in Right of Canada, as represented by the Minister of Industry, 2023

# The Social Data Linkage Environment (SDLE)



# Social Data Linkage Environment (SDLE)

Coverage: historical database of Canadians built by combining administrative files

Over 160 data sources already linked

Facilitates the creation of linked population data sets for social analysis

## Analytical Opportunities:

- Use existing administrative and survey data to address important research questions and inform socioeconomic policy
- Generate new information and valuable insights without the need for any additional data collection



# Enhance your data with SDLE

Reduce costs of research projects.

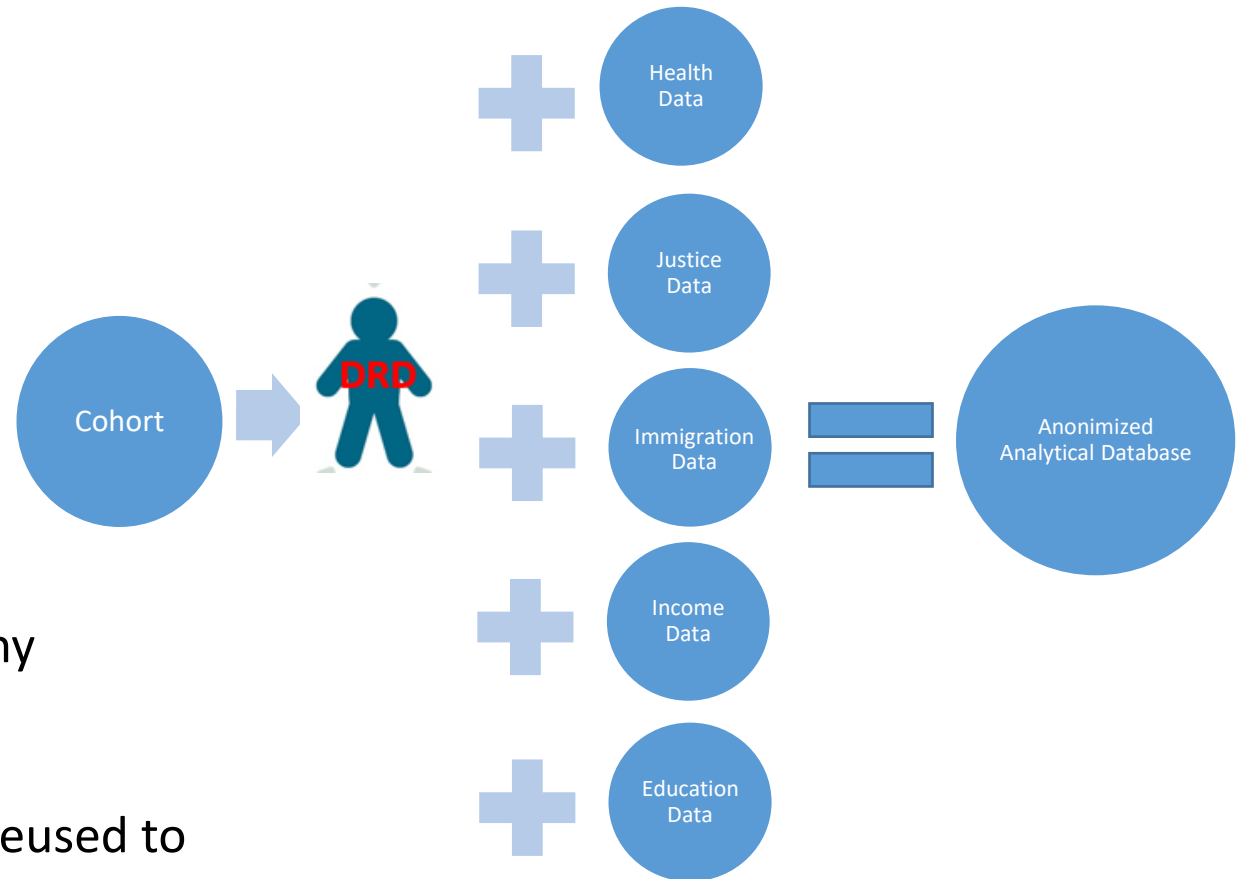
Standardised approach to record linkage

Reduce response burden.

Using existing administrative and survey data fills data gap to address important research questions and inform socioeconomic policy.

Generate new information without the need for any additional data collection.

SDLE files are linked once, and linkage results are reused to create analysis files.





# SDLE in Statistics Canada Data

**Longitudinal Immigration  
Database (IMDB)**

**Education and Labour Market  
Longitudinal Platform (ELMLP)**

**Canadian Census Health and  
Environment cohorts (CanCHECs)**

**Criminal Justice Relational  
Database (CJRD)**

**Census of Population**

**Survey of Financial Security (SFS)**

**Canadian Community Health Survey (CCHS)**

**General Social Survey (GSS)**

**Canadian Social Survey (CSS)**

[DRD linkage status - Social Data Linkage Environment \(SDLE\) - Environnement de couplage de données sociales \(ECDS\) - Statcan Confluence](#)



# Linkage Projects in collaboration with SDLE

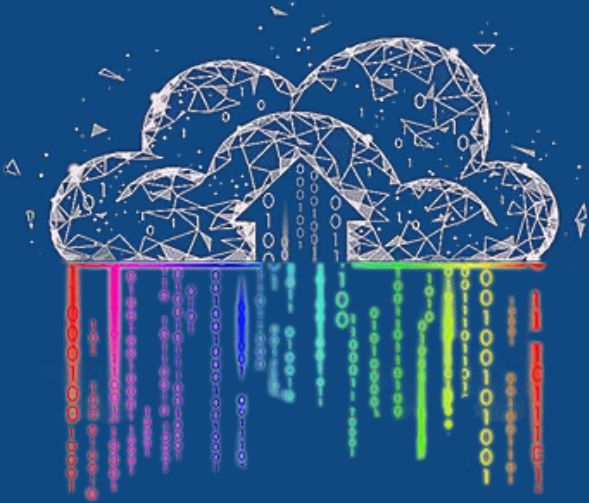
## Justice Outcomes

- Profiles of repeated contact with the multiple provincial criminal justice systems (ON, SASK, etc)
- Examination of the long-term community adjustment of offenders
- Ontario Bail and Remand Project

## Immigration and Health Outcomes

- Canadian Cancer Treatment Linkage Project
- Linkage of the Canadian Community Health Survey to the Longitudinal Immigration Database
- Linkage of the Discharge Abstract Database, National Ambulatory Care Reporting System, and the Ontario Mental Health Reporting System to the Canadian Vital Statistics Death Database, 2000-2021 (ongoing)
- **Results observed for specific populations**
- British Columbia Opioid Study for the Municipality of Surrey
- Canadian Armed Forces Cancer and Mortality Study
- **Observed educational outcomes**
- Canada Student Loans and Learning Loans Program





# The Linkable Open Data Environment (LODE)



# Linkable Open Data Environment (LODE)

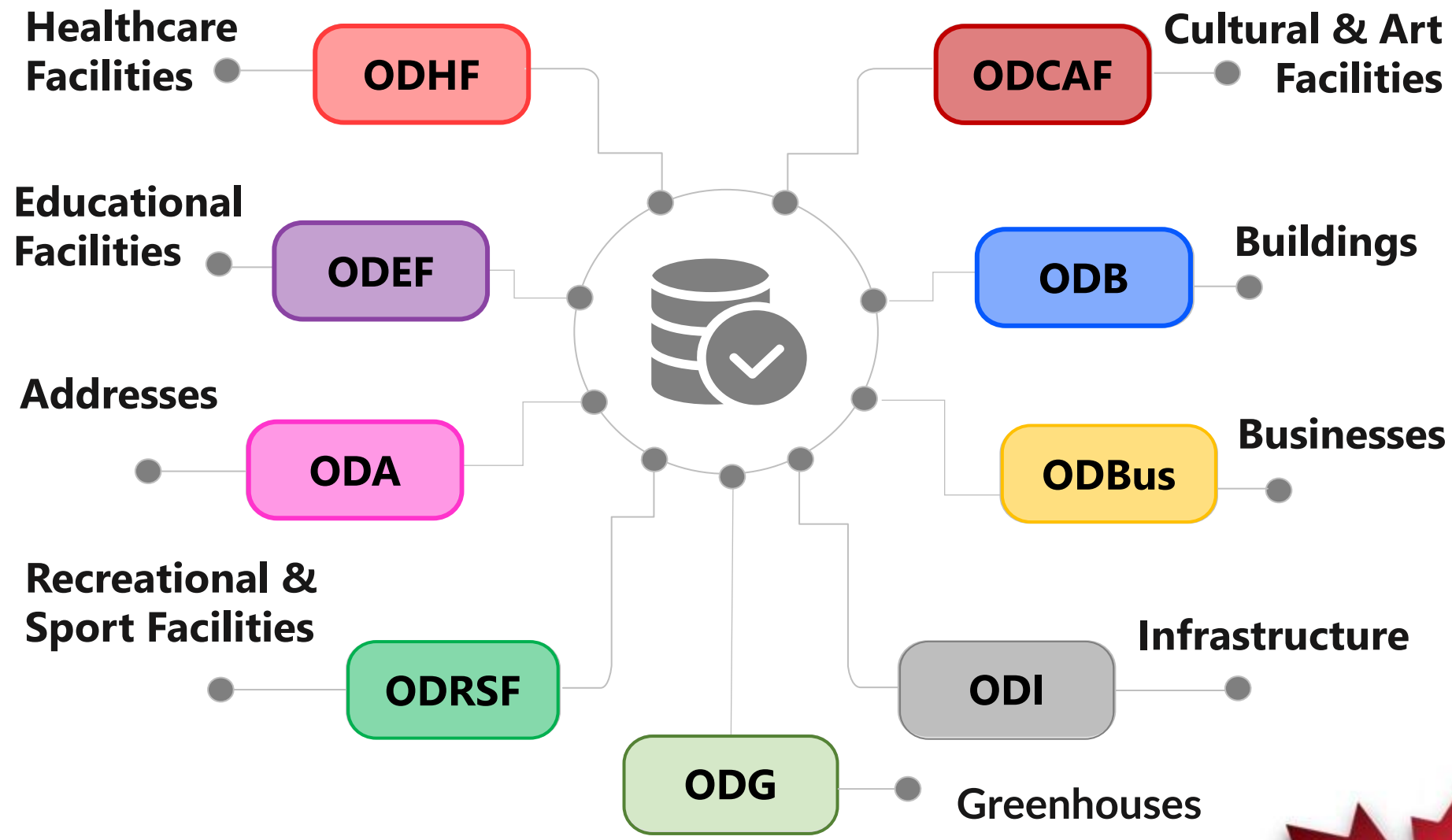
Coverage varies by subject matter

100's of open data sources

Facilitates open research on public and private facilities across Canada by harmonizing data from all levels of government and other public facing entities.

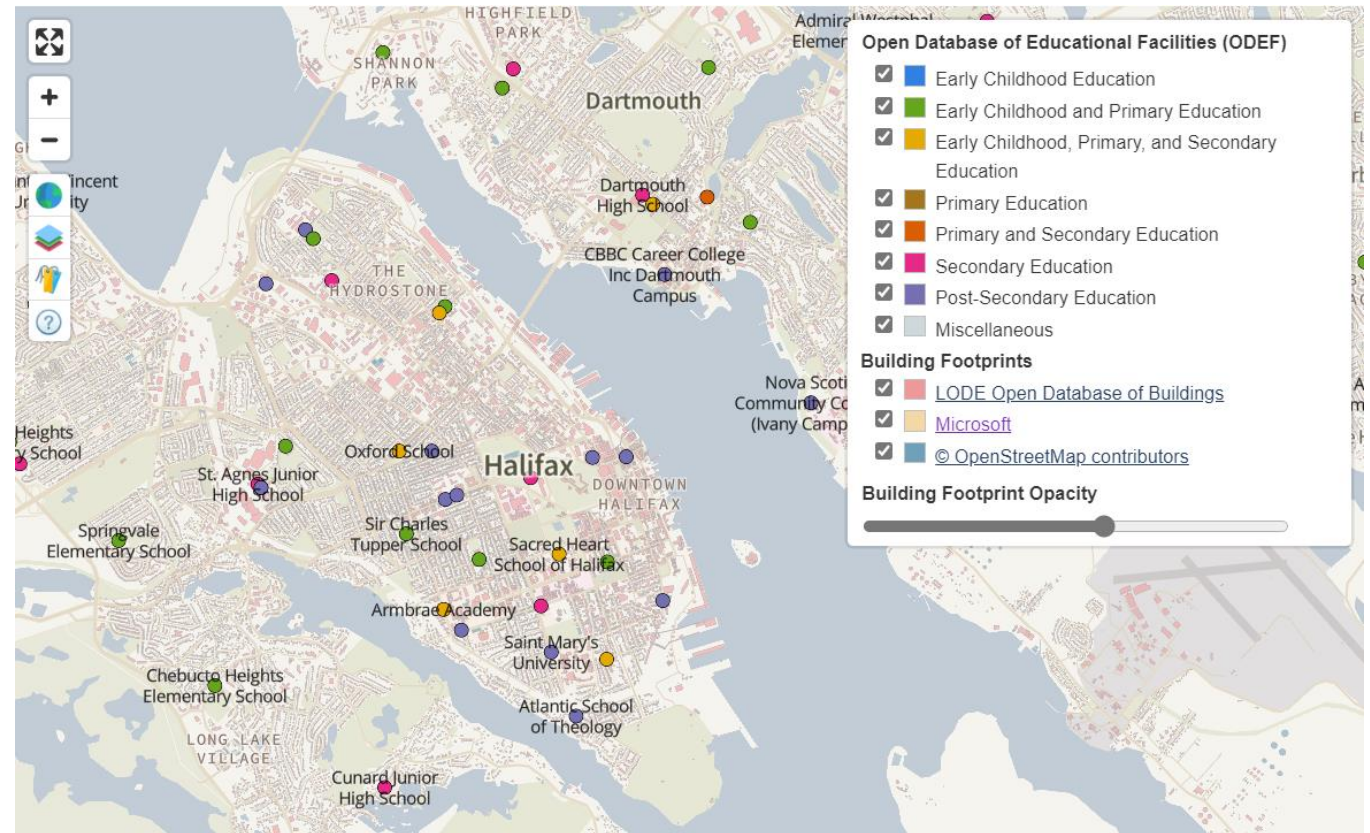
Analytical Opportunities:

- Open Government Licence allows easy sharing and collaboration
- Contributes to flagship products such as the Spatial Access Measures



# Open data and more

- Open-source code
  - Data processing with Python
  - Transparent metadata
- Open collaborations
  - Open Addresses
  - Outreach
- Open data viewer



[Linkable Open Data Environment Viewer](#)

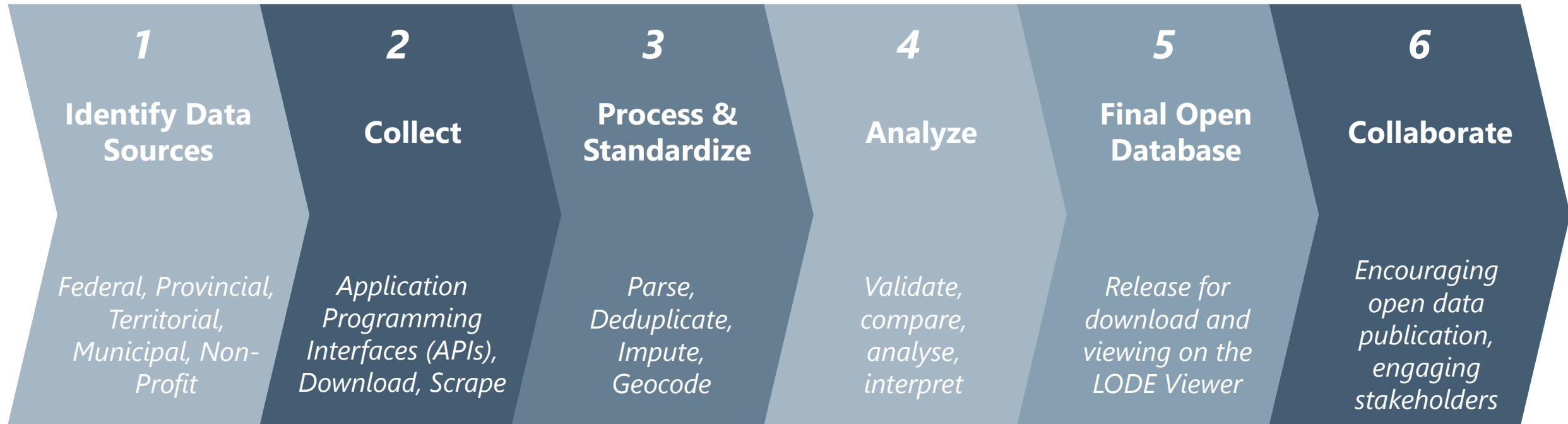
# An example – The Open Database of Infrastructure

- A collection of open data containing the location of a selection of infrastructure across Canada
- Made available under the [Open Government Licence - Canada](#)
- Contains:
  - Bridges and tunnels
  - Potable water
  - Stormwater and wastewater
  - Solid waste
  - Pedestrian and cycling paths



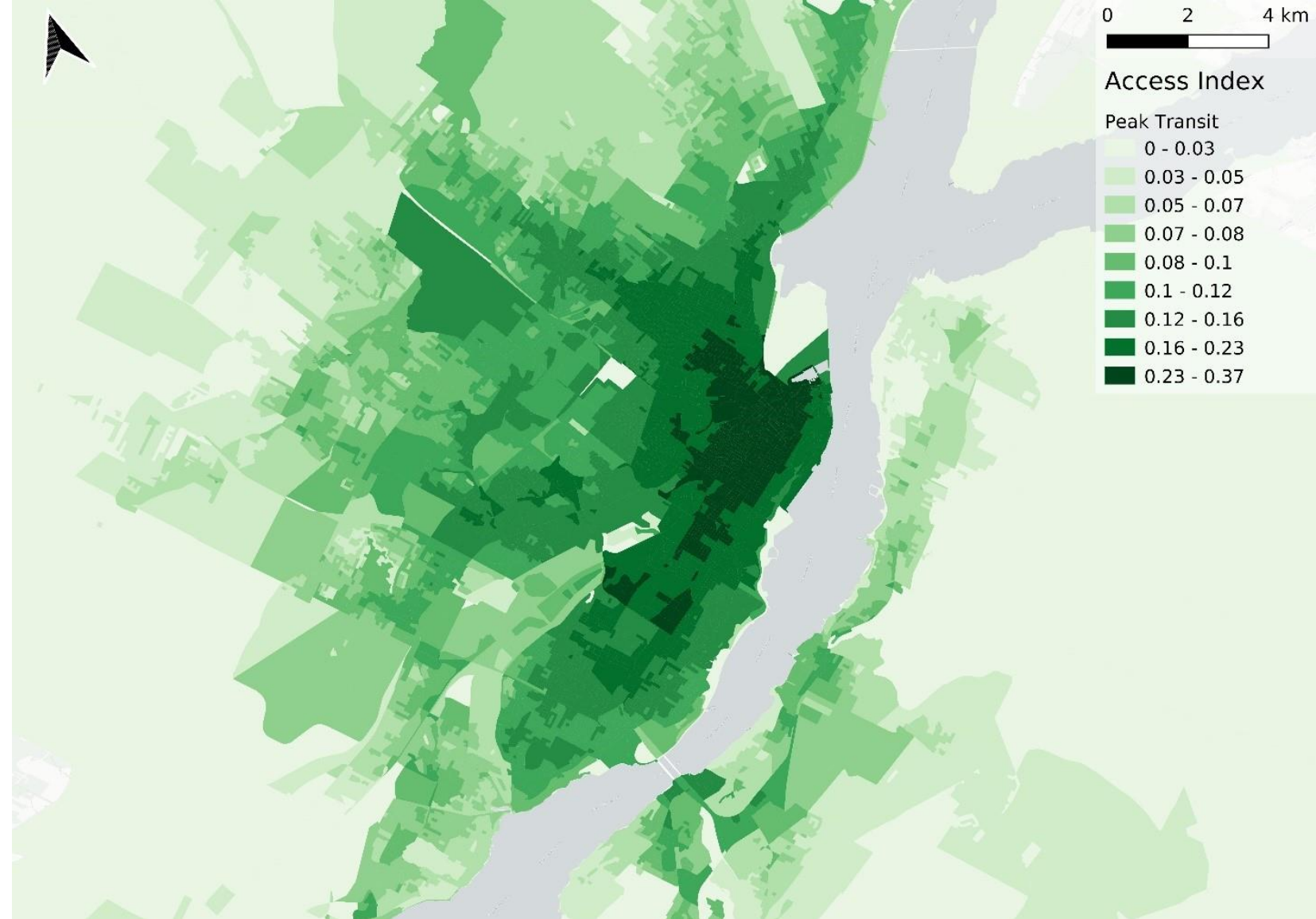
[Linkable Open Data Environment Viewer](#)

# LODE Process



# How is the LODE used?

- Each LODE product can be used to feed into subject matter analysis
- For example, the [Spatial Access Measures](#) explore access to seven amenities and services across Canada by four modes of active and public transportation
- The Open Database of Educational Facilities was used to compute access to post-secondary education via public transit



Access to post-secondary education, peak transit hours, ville de Québec



# Iterative development and improvement in the LODE

- Expanding subject matter:
- November 2023: Release of the Open Database of Businesses
- May 2023: Release of the Open Database of Infrastructure
- February 2023: Release of the Open Database of Greenhouses
- Adding new variables:
- Spring 2024: Upgrade of the Open Database of Educational Facilities
  - Identification of new Official Language Minority Schools (OLMS)
  - Addition of French immersion schools
- Spring 2024: Open Database of Infrastructure V 2.0

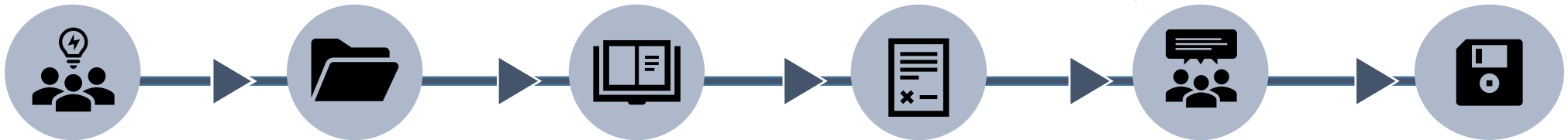


# Data Linkage Access Process



### Cost Recovery

## Process to Initiate a Linkable Project



Meet with Statistics Canada to discuss your information needs

Data Dictionaries to identify variables of interest

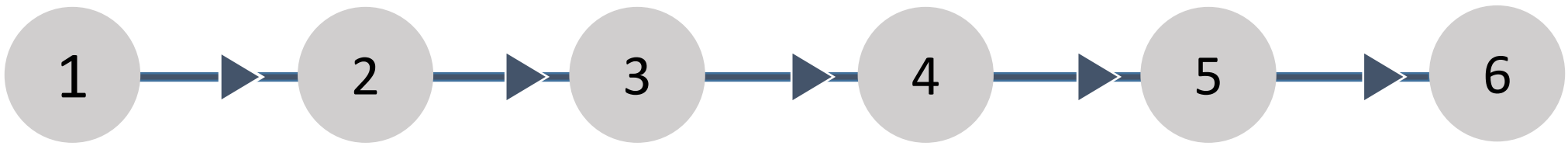
Project Proposal Process

Sign Letter of Agreement

Work with Statistics Canada throughout the project cycle

Tables and Reports Transferred

### Data Access Division (DAD)



1 Create a detailed research proposal

2 Submit application via Micro Data Access Portal

3 If approved, applicants must become deemed employees

4 Required access to files is granted

5 Work with Statistics Canada throughout the project cycle

6 Analysis and Output vetted by Statistics Canada before releasing

# Thank you! Please stay connected.

Questions?

**Julio Rosa, Chief**

Business Linkable File Environment (BLFE)

613-796-4371 [julio.rosa@statcan.gc.ca](mailto:julio.rosa@statcan.gc.ca)

Contact us:

**Chantal Poirier, Chief**

Social Data Linkage Environment (SDLE)

613-854-8237 [chantal.poirier@statcan.gc.ca](mailto:chantal.poirier@statcan.gc.ca)

**Bjenk Ellefsen, Chief**

Linkable Open Data Environment (LODE)

(343) 573-2370 [Bjenk.Ellefsen@statcan.gc.ca](mailto:Bjenk.Ellefsen@statcan.gc.ca)



# ANNEX 1

## Administrative data sources (B\_LFE)

<https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=6000>

- Business Register (BR) - 1996 to 2022 (SDDS 1105)
- Longitudinal Employment Analysis Program (LEAP) - 2000 to 2019 (SDDS 8013)
- General Index of Financial Information (GIFI - T1) unincorporated businesses - 2005 to 2021
- General Index of Financial Information (GIFI-T2) incorporated businesses - 2000 to 2021
- T4 (Statement of Remuneration Paid) Supplemental File - 1997 to 2021
- PMF (Personal Master File) of primary owner/manager of enterprise (including some variables from Census 2006 to 2021 [SSDS 3901] and the Longitudinal Immigration Database 2000 to 2021 [SSDS 5057]) - 2007 to 2014
- TED (T1 Enhanced Database) of primary owner/manager of enterprise (including some variables from Census 2006 to 2021 [SSDS 3901] and the Longitudinal Immigration Database 2000 to 2021 [SSDS 5057]) - 2012 to 2021
- Payroll Deductions Account (PD7) - 2001 to 2021
- Exporter Register - 2010 to 2021 (SDDS 2201)
- Importer Register - 2012 to 2021 (SDDS 2201)
- Patents (Canadian Intellectual Property office) - 2001 to 2006
- United States Patent Office (USPTO) Canadian enterprises only - 2000 to 2011
- Business Innovation and Growth Support (BIGS) - 2007-2008 to 2020-2021 (SDDS 5304)
- Agriculture Tax Data Program (ATDP) - 2016 to 2019 (SDDS 3447)
- Charities - 2003 to 2020
- Canadian Patent Research Database (CPRD) - 2001 to 2015
- Diversity and Skills Database (DSD) - 2001 to 2019 (SDDS 5228)



# ANNEX 2

Survey data sources (B\_LFE)

<https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=6000>

- Annual Survey of Research and Development in Canadian Industry (RDCI) - 2000 to 2020 (SDDS 4201)
- Balance of Payments: Canadian Direct Investment Abroad (CDIA) 2000 to 2013 (SDDS 1537)
- Balance of Payments: Foreign Direct Investment in Canada (FDIC) - 2000 to 2013 (SDDS 1537)
- Trade in Commercial Services (TICS/TIS) 2000 to 2014 (SDDS 1536)
- Survey of Innovation and Business Strategy (SIBS) 2009, 2012, 2017 (SDDS 5171)
- Surveys of Innovation (INNO) 2003, 2005 (SDDS 4218)
- Survey of Electronic Commerce Technology (SECT) 2000 to 2007 (SDDS 4432)
- Survey of Advanced Technology (SAT/ATS) 2007, 2014 (SDDS 4223)
- Survey of Commercialization of Innovation (COI) - 2007 (SDDS 5140)
- Survey on Financing (and Growth) of Small and Medium Enterprises (SFSME/SFGSME) - 2004, 2007, 2011, 2014, 2017, 2020 (SDDS 2941)
- Survey of Intellectual Property Management (SIPM) - 2010 (SDDS 5183)
- Survey of Digital Technology and Internet Use (SDTIU) - 2012, 2013, 2019, 2021 (SDDS 4225)
- Survey of Regulatory Compliance Cost (RCC) - 2011, 2016 (SDDS 5093)
- Schedule 32 - Claim for Scientific Research and Experimental Development (SR&ED) Carried out in Canada (Form T661) - 2000 to 2021
- Survey of Intellectual Property Awareness and Use (IPAU) - 2019 (SDDS 5291)



# Types of Linkages

## Deterministic or exact match:

- Candidate record pairs are compared using a set of comparison functions that only require exact similarity. Usually, this type of straightforward match is based on **unique identifiers** (Business number; SIN number...)

## Probabilistic match:

- This approach calculates match weights based on the similarity (Names; Addresses; Postal codes...).

