



# CDER: An Overview and Recent Development on Business Microdata

Jiang Beryl Li

Economic Analysis Division, Statistics Canada



Delivering insight through data for a better Canada



Statistics  
Canada

Statistique  
Canada

Canada

# Roadmap

- ❑ What is CDER?
- ❑ What microdata are available?
- ❑ What is new?
- ❑ How to access CDER?

## Business Microdata: challenges

- ❑ Identity of units must be protected
  - No individual or business can be directly or indirectly identified
- ❑ Unlike social microdata, units in business microdata are readily identifiable even without direct identifiers
  - The distribution of business data is often highly skewed (i.e., it is difficult to protect the identity of large firms)
  - The sparsity of the data in certain dimensions (e.g., detailed industry and geography)
- ❑ There are greater financial incentives to identifying the particulars of some firms and their competitors

# Canadian Centre for Data Development and Economic Research (CDER)

- ❑ CDER was created in 2011 to make better use of Statistics Canada's business microdata holdings without compromising security
  - Serves as a repository for business microdata
  - Leads the development of new business microdata
  - Provides external researchers with secure access for research-oriented projects
  - Five safes of access: projects; researchers; workplace; data and outputs
- ❑ Institutions served
  - About 30 Canadian and foreign academic institutions
  - 16 federal government departments and agencies, provincial government and non-profit institutions
- ❑ Until recently, access was only provided at Statistics Canada's head office in Ottawa
  - Application and approval process required for access
  - Researchers must cover the full cost of their project

## Business Microdata at CDER

- ❑ Stand-alone, research-ready data already in use
  - E.g., Trade by Characteristics of Enterprises; Survey of Financing and Growth of SMEs linked to tax data; and Annual Survey of Manufactures
- ❑ [Linkable File Environment \(LFE\)](#)
  - Specific variables from a set of files where linkages have been done, but files are so large that extractions are made upon request
- ❑ Developmental datasets and other linkage environments
  - Analytical databases containing derived variables for specific analyses (e.g., National Accounts Longitudinal Microdata File); additions to LFE; other linkage environments (e.g., Canadian Employer-Employee Dynamic Database); new stand-alone data



## Linkable File Environment (LFE)

- ❑ The LFE is an environment that contains datasets from administrative and surveyed sources that are linkable (the links have been done, proven and documented), but because of the size of the databases involved are not stored as one database

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

- ❑ Statistics Canada's Business Register is the "central source" of the LFE environment
- ❑ Variables are extracted and "custom research dataset" are produced for specific analytical project



## Administrative Datasets in the LFE

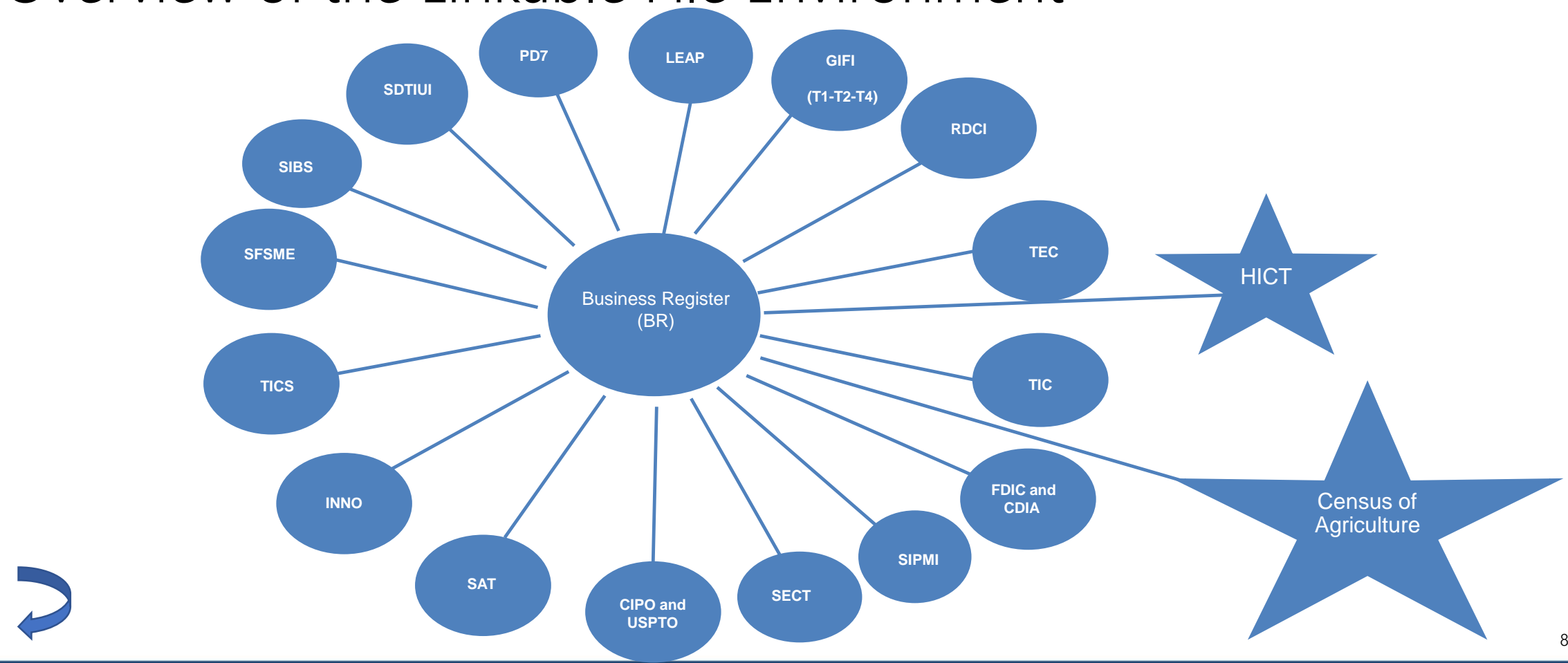
- ❑ Business Register (BR), 2000-2017
- ❑ Longitudinal Employment Analysis Program (LEAP), 2000-2017
- ❑ General Index of Financial Information (GIFI – T1), 2005-2016
- ❑ General Index of Financial Information (GIFI – T2), 2000-2017
- ❑ Statement of Remuneration Paid (GIFI – T4), 2000-2016
- ❑ Payroll Deduction Accounts (PD7), 2001-2017
- ❑ Trade by Exporter and Importer Characteristics (TEC and TIC), 2010-2018
- ❑ Patents (Canadian Intellectual Property Office), 2001-2006
- ❑ United States Patent Office (USPTO – Canadian Enterprises only), 2000-2011
- ❑ Horizontal Innovation and Clean Tech (HICT), 2007-2016

## Survey Datasets in the LFE

- ❑ Survey of Electronic Commerce and Technology (SECT), 2000-2007
- ❑ Survey of Innovation (INNO), 2003 and 2005
- ❑ Survey of Innovation and Business Strategy (SIBS), 2009, 2012, and 2016 (Summer 2019)
- ❑ Survey of Advanced Technology (SAT), 2007 and 2014
- ❑ Survey of Intellectual Property Management (SIPM), 2010
- ❑ Survey of Financing and Growth of SMEs (SFGSME), 2007, 2011, 2014
- ❑ Survey of Digital Technology and Internet Use (SDTIU), 2012 and 2013
- ❑ Annual Survey of Research and Development in Canadian Industry (RDCI), 2000-2013
- ❑ Foreign Direct Investment and Canadian Direct Investment Abroad (FDI and CDIA):2000-2017
- ❑ Trade in Commercial Services (TICS), 2000-2014
- ❑ Census of Agriculture, 2016



# Overview of the Linkable File Environment



# Trade by Enterprise Characteristics

- ❑ Enterprise level, 2010-2018
- ❑ By Industry, province, CMA, concentration and by employment of enterprise
- ❑ Trade by Exporter Characteristics (TEC)
  - Value of exports by HS8 commodity
  - By country of destination, including U.S. States breakdown
- ❑ Trade by Importer Characteristics (TIC)
  - Value of imports by HS10 commodity
  - By country of origin, country of export, and by related-parties
- ❑ Uses
  - Trade dynamics: barriers to trade, firms' life cycle in the export/import market
  - Firm dynamics: trade orientation, owner demographics and firm characteristics (if linked to administrative data)
  - Industry or product concentration: Herfindahl index

# Survey of Financing and Growth of SMEs (SFGSME)

- ❑ Cross-sectional survey
  - 2000, 2001, 2004, 2007, 2011, 2014 and 2017
- ❑ Survey information
  - Use of financing during start up
  - Requests for financing (term loans, mortgages, lines of credit, credit cards, government loans, equity) and outcomes (approved/rejected, collateral, term, interest rate, amounts requested/received)
  - Business information (exports, R&D, innovation, IP use, plans for growth, public procurement participation)
  - Owner information (age, education, experience, country of birth, language of primary decision maker; %female, %aboriginal, %visible minority)
- ❑ Linked to admin. data on firm performance, 2000-2017
- ❑ Uses
  - Financing cost; lending relations; and financial constraints on firm dynamics

# Annual Survey of Manufactures (ASM)

- ❑ ASM (1961-2012) – a series of longitudinal datasets
  - Cross-sectional, establishment level survey of manufacturing
  - Principal industrial statistics and commodity file...100s of variables in the latest database
  - Essentially a census up to 2012
- ❑ ASM has been linked to other sources
  - National Pollutant Release Inventory (NPRI) and Greenhouse Gas Reporting Protocol (GHGRP), plant level, 2000 to 2012
  - General Index of Financial Information (GIFI), ASM-enterprise level, 2000 to 2012
  - Research and Development in Canadian Industries, ASM-enterprise level, 2000 to 2009
- ❑ Uses
  - Productivity, trade, and innovation

# Longitudinal Employment Analysis Program (LEAP)

- ❑ Firm level, 1983 to 2017
- ❑ Program information
  - Employment, payroll, first/last year, industry and province
  - Labour tracking
- ❑ T2-LEAP
  - LEAP linked to core administrative data (including capital investment program) from the corporate tax system covering 1983 to 2017; 1997, 2004, 2007, 2008 to 2017 vintages
- ❑ Uses
  - Firm entry and exit
  - Job creation and destruction

# National Accounts Longitudinal Microdata File (NALMF)

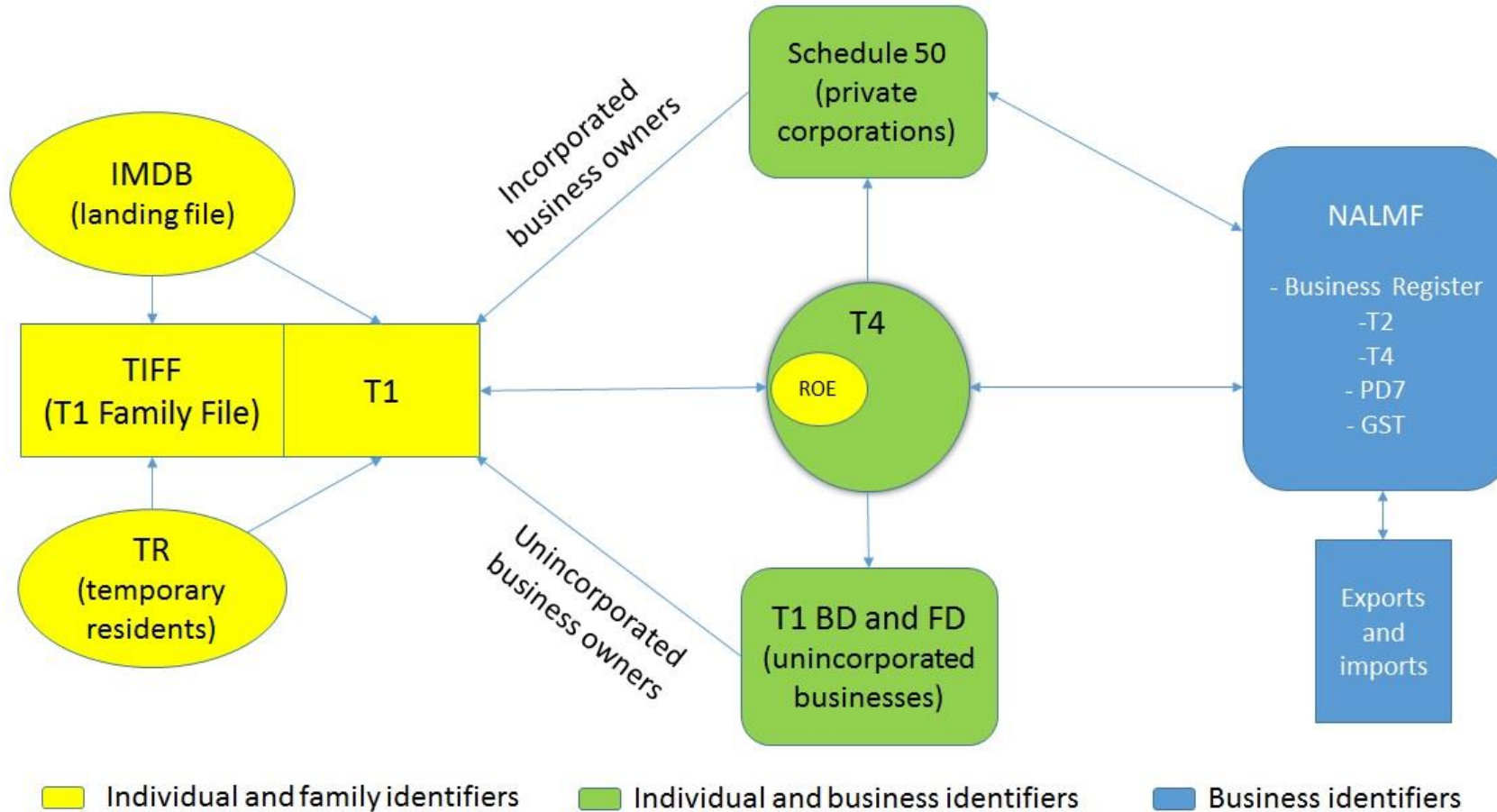
- ❑ Longitudinal database of Canadian enterprises, 2000-2016
  - Successor to T2-Longitudinal Employment Analysis Program
  - Tracks a richer set of firm characteristics over time (e.g., employment, payroll, revenue, profit, assets, tangible assets, R&D, investment, and productivity)
  - Updated longitudinal structure
- ❑ Main data sources
  - Statistics Canada's Business Register; Corporation Income Tax (T2); Payroll Account Deductions (PD7) and Statements of Remuneration Paid (T4 slip); and Goods and Services Tax (GST/HST)
- ❑ Uses
  - GDP and employment across firm-size categories
  - Activities at sub-provincial levels (e.g., CMAs and ERs)
  - Firm dynamics, survival, and growth



# Canadian Employer-Employee Dynamics Database (CEEDD)

- ❑ Matched employer-employee database, 2001 to 2016
- ❑ Covers the universe of
  - Individual tax filers and their families
  - Employees receiving T4 slips
  - Unincorporated businesses and corporations
  - Unincorporated business owners reporting self-employment income
  - Owners of Canadian-controlled Private Corporations
- ❑ Links across various administrative tax files
  - T1: Personal Master File (T1PMF); Historical Personal Master File (T1H); Family File (T1FF)
  - Longitudinal Immigration Database (IMDB); Temporary Residents File (TR)
  - T1: Business Declaration File (T1BD); Financial Declaration File (T1FD)
  - National Accounts Longitudinal Microdata File (NALMF)
  - T2 Schedule 50; T4 Statement of Remuneration Paid Files (T4); Record of Employment (ROE)
  - Trade by Exporter Characteristics (TEC); Trade by Importer Characteristics (TIC)

# CEEDD – Linkage environment



## Possible Uses of CEEDD

### ❑ Labour market and firm dynamics

- E.g., productivity and wage dispersion; the impact of merger and acquisition; the impact of start-ups; entrepreneurs' hiring decision; and entrepreneurship on a fine spatial scale, etc.

### ❑ Immigrants

- E.g., careers and network; initial firm allocation and earning growth; business ownership; financial constraints and survivals of immigrant-owned firms; and trade orientation, etc.

### ❑ Government programs: evaluation and policy development

- E.g., employment at the economic regional level; business ownership by gender, age, immigration status, and by marital status

# CEEDD

## ❑ Business Owner Module (2005-2016)

- Immigrant
- Gender
- Age

## ❑ Geographic information (2011 census boundary)

- Census metropolitan areas (CMAs)
- Self-contained labour areas (SLAs)

## ❑ Linkage to the CEEDD (assessment to be completed in March 2020)

- Census: Indigenous group, occupation, and education, etc.
- PATSTAT: Applications, grants, patent families, citations, patent classifications, applicants, inventors, and legal status

## Employment and Social Development Canada (ESDC): Request for Proposals (Apply now!)

- ❑ A set of linked files to CEEDD to assess the labour market impacts of the Temporary Foreign Worker (TFW) Program
  - ESDC: Record of Employment File; TFW Program File (2010+)
  - IRCC: Temporary Resident File; Landed Immigrant File
- ❑ ESDC is supporting effective use of these data by issuing up to two contracts to develop a research network that will address:
  - The impacts of the TFW Program on the Canadian labour market, including potential suppression of wages and displacement of Canadian workers;
  - The extent to which the TFW Program is meeting the needs of Canadian firms and the labour market as a whole; and,
  - The effects of the reformed TFW Program on Canadian businesses and the labour market.
- ❑ Apply at <https://buyandsell.gc.ca/tma/node/875503>
  - Closing date is September 30, 2019

# New Business Microdata at CDER

## ■ International Accounts and Trade

- Canada and the World Statistics Hub (13-609X): Updated on an ongoing basis
- Trade by Enterprises Characteristics: 2010-2018 (May 2019)
- Trade by Enterprises Characteristics, by census metropolitan area: 2010-2018 (Sep 2019)
- Foreign Direct Investment and Canadian Direct Investment Aboard: 2000-2018 (Apr 2019)
- Outward Foreign Affiliates Statistics (FATS): 2011-2016 (RY2017 in fall 2019)
- Enhanced details on securities statistics: 2015Q1 to 2019Q1
- Inward FATS and activities of multinational firms in Canada (RY2010-2017 in Nov 2019)
- Canada's international trade in services: 2017 (Oct 2018)



## New Business Microdata at CDER (cont'd)

### □ Investment, Science, and Technology

- Canadian Survey of Cyber Security and Cybercrime (RY2017)
- Survey of Innovation and Business Strategy (RY2017 in Jun 2019)
  - Innovation
  - Advanced technology use
  - Business strategies and practices
  - Global value chains
- Canadian Housing Statistics Program (RY2017-2018)
  - Link numerous administrative files (e.g., tax files, property assessment files, etc.)
  - Used to understand and address foreign-home-ownership-related issues
  - By geography(CMAs & CSDs in Vancouver and Toronto), property/owner characteristics, property use/financing

## New Business Microdata at CDER (cont'd)

### □ Farm-level data

- Census of Agriculture: 1986, 1991, 1996, 2001, 2006, 2011 and 2016 (longitudinal files underlined)
- Farm Financial Survey, 2013, 2015 and 2017
- Livestock, crops and horticulture (potential)

### □ Price Indices

- Cover a range of goods, services and construction outputs
- By type: transaction price; unit value; margin price; list price; and model price
- Unpublished indices are available in many cases (incl. by geography/product/industry) and with a more recent starting date

# Application Process

- ❑ By type of researcher: government (federal/provincial) and non-government
- ❑ For non-government researchers, the process is as follows
  - 1) Contact CDER and draft a proposal
  - 2) Submit final proposal
  - 3) Proposal evaluation: peer review and approved by an internal Statistics Canada committee
  - 4) Undergo a security screening
  - 5) Take the Oath of the Office of the Statistics Act to become deemed employee
  - 6) Sign a Microdata Research Contract
- ❑ Contact:
  - Website: <http://www.statcan.gc.ca/cder>
  - Email: [statcan.cder-cdre.statcan@canada.ca](mailto:statcan.cder-cdre.statcan@canada.ca)

# Improving Access to Business Microdata

## 1 – Pilot Projects

- ❑ T2-Longitudinal Employment Program (2001-2015): April 2018
- ❑ Survey of Financing and Growth of Small and Medium-sized Enterprises 2014 linked to administrative data (2010-2014): November 2018
- ❑ Project setup
  - Calls for proposal and webinars
  - Approved research projects that focus on multivariate analysis are being carried out in the RDCs in 2018-2019 and 2019-2020
  - Researchers develop their models using the synthetic/treated data at the RDCs
  - Programs are then sent to CDER by RDC analysts to be run on the actual data
  - Final results are released from CDER by e-mail
- ❑ Ultimate objectives
  - From the experience gained, similar datasets can be placed in the RDCs

# Improving Access to Business Microdata (Cont'd)

## 2 – RDC Releases

- ❑ Core business microdata (March 2020)
  - An LFE extraction: A selection of administrative and survey data
  - The variable list is to be finalized after consultations with internal/external stakeholders
- ❑ Core Canadian Employer-Employee Dynamic Database (March 2021)
  - Stay tuned!

# Improving Access to Business Microdata (Cont'd)

## 3 – Productivity Partnership

- ❑ Team of experts from the academic, private, and public sectors
- ❑ Lowers barriers to access business data for academics and students by providing funding for productivity research projects
- ❑ Statistics Canada liaison researcher: Beryl Li, JiangBeryl.Li@canada.ca
- ❑ Important partner in the development of synthetic datasets
- ❑ Collaboration and support to clarify the CDER application process steps

<https://productivitypartnership.ca>



# Thank You!

For more information, please visit  
[www.statcan.gc.ca/eng/cder/index](http://www.statcan.gc.ca/eng/cder/index)

