

Canadian Centre for Data development and Economic Research (CDER)

Lydia Couture (Economic Analysis Division)

December 16, 2019

Delivering insight through data for a better Canada



Statistics
Canada

Statistique
Canada



Canada

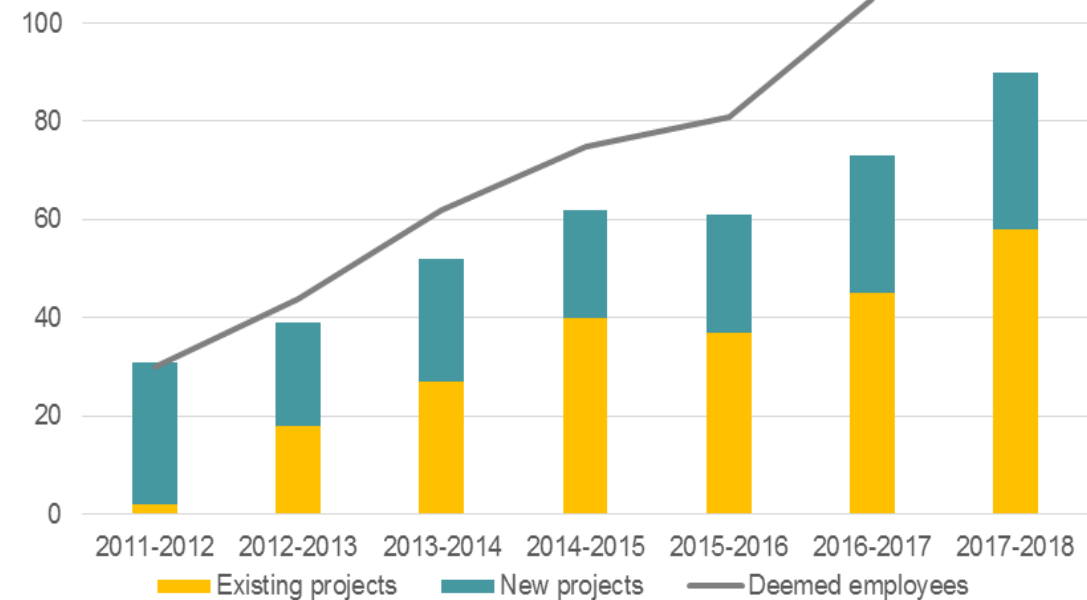
Overview

1. What is the Canadian Centre for Data Development and Economic Research (CDER)?
2. Application process
3. Data available
4. Initiatives to improve access



Context

- CDER was established in 2011 at Statistics Canada's headquarters to provide secure access to business microdata for approved analytical research projects
- Institutions served:
 - About 30 Canadian and foreign academic institutions
 - 16 federal government departments and agencies, provincial governments and non-profit institutions
- Provides researchers with secure access to business microdata for *research-oriented projects* that serve the mandate of Statistics Canada



Challenges with business microdata

- **Identity of units (individuals or enterprises) must always be protected**

Challenge: The units in business microdata are more easily identifiable than in social data even without direct identifiers

- **The distribution of firms is often skewed**

- It is difficult to protect the identity of large firms
- Data are sparse even in high-levels of geography or highly aggregated industries

Challenge: Confidentiality vetting rules specific to business data

There are financial incentives to re-identifying businesses and using the data for non-statistical purpose

Challenge: The conflict of interest issues that need to be resolved in regards to business data are different from the ones for social data

Key information

- CDER approval process is similar to that of the RDCs
- Until recently, access was provided only at Statistics Canada's headquarters in Ottawa
- Researchers must cover the full cost of their project

CDER Application Process

Step 1 - Contact CDER and draft a proposal

- a) The justification for the research:
 - context;
 - the research question;
 - contribution to the literature.

- b) The analytical framework and the data requirements:
 - detailed data requirements;
 - proposed methodology;
 - justification for using micro data;
 - expected outputs;
 - software requirements;
 - expected length of project.

Step 2 - Submit project proposal

- a) Application for accreditation:
 - CV that demonstrates experience and technical competence;

- b) A letter from lead researcher indicating:
 - how project costs will be covered;
 - their ability to abide all the terms and conditions of becoming a deemed employee and the conditions in the research contract;
 - that they have no conflicts of interest to declare;
 - that they can commit to producing a research paper for Statistics Canada.

Step 3 - Evaluation of proposal

1. Peer review
 - a) Project is being funded by SSHRC: Statistics Canada will take this as recommendation that the project has been peer reviewed and that researchers are qualified
 - b) Require Statistics Canada to conduct a peer review: Statistics Canada will solicit two reviews from an external panel of experts at a cost of \$200 that will be paid by applicants

2. An internal Statistics Canada committee will review to ensure projects falls under the mandate of Statistics Canada

Step 4 - Complete the security screening process

Step 5 - Sworn in as deemed employees of Statistic Canada and sign a micro data research contract

Data available at CDER

- 1) Stand-alone, research-ready data already in use
- 2) Linkable File Environment (LFE)
 - Specific variables from a set of files where linkages have been done, so large that extractions are made upon request
- 3) Developmental datasets and other linkage environments
 - Analytical databases containing
 - derived variables for specific analyses
 - additions to LFE
 - other linkage environments
 - new stand-alone data

Stand-alone databases

Survey of Financing and Growth of SMEs

- Cross-sectional survey in 2000, 2001, 2004, 2007, 2011, 2014, and 2017
- Linked to administrative data on firm performance before and after survey years, 2000 to 2017
- Themes:
 - 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)

Survey of Innovation and Business Strategies (SIBS)

- Cross-sectional survey in 2009, 2012, 2017
- Can be linked to administrative data through the Linkable File Environment
- Research projects linking the common respondents in the surveys have been approved

Content

- Strategic decisions – locations, outsourcing, global value chain participation, lost-cost/product differentiation,
- Innovation activities – advanced technology use, product/process/marketing/organizational innovation, obstacles
- Operational tactics – production and human resource management, business practices, relationship with suppliers

Trade by Exporter/Importer Characteristics (TEC & TEC)

- Enterprise-level, 2010 to 2018
- Industry, province, CMA, employment of enterprise
- Value by HS-8/HS-10 commodity codes and country of destination/origin

Longitudinal Employment Analysis Program (LEAP)

- Firm entry and exit, job creation and job destruction, and payroll
- Labour tracking
- 2001 to 2016 vintages covering 1983 to 2016
- T2-LEAP – LEAP linked to core administrative data (including capital investment program) from the corporate tax system covering 1983 to 2016; 1997, 2004, 2007, 2008 to 2016 vintages

Other databases

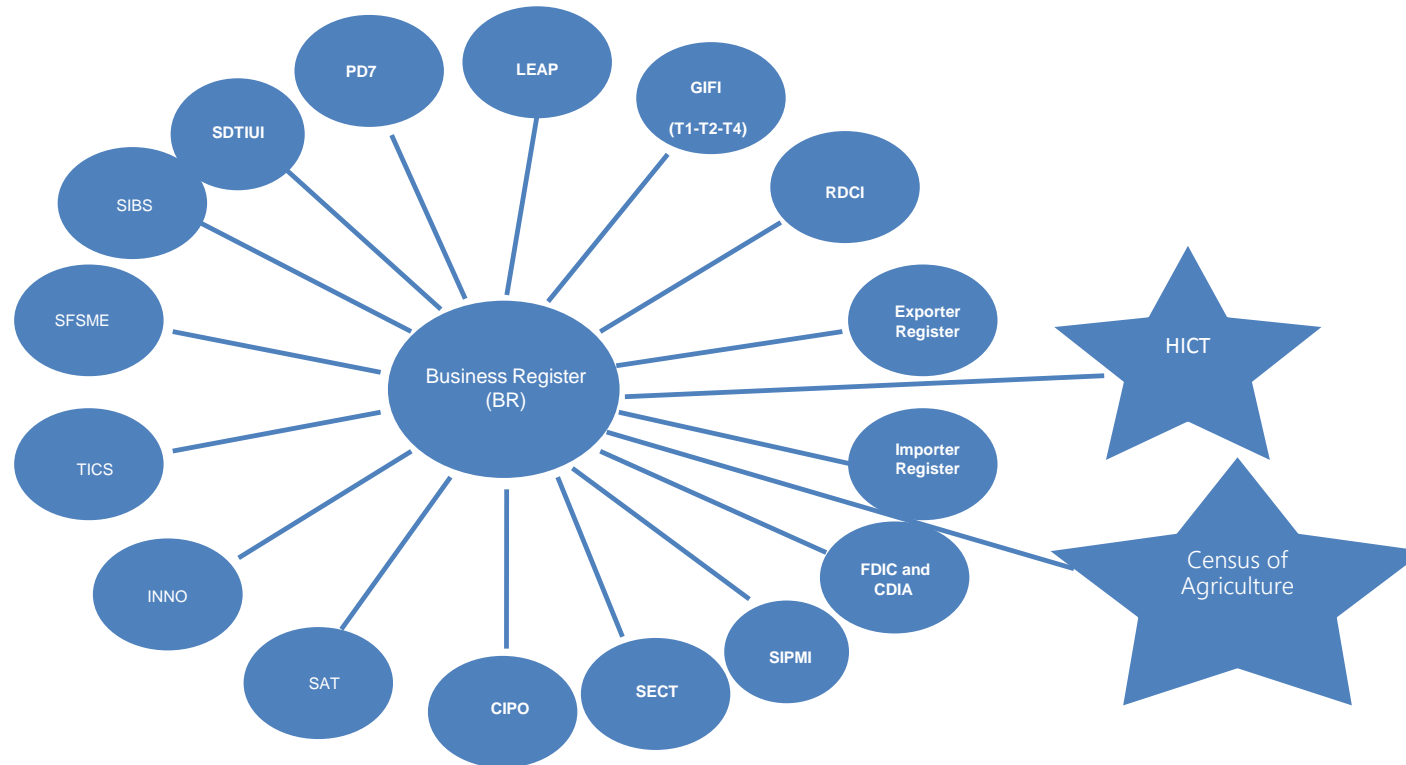
- Retail Trade Survey
- Wholesale Trade Survey
- Consumer Price Research Database
- Producer Price Database
- Bankruptcy Database

Linkable File Environment (LFE)

Linkable File Environment

- 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.
- The LFE has data on the business registry; corporate taxes; pay deductions; export and import data as well as many business surveys.
- <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

Overview of the Linkable File Environment



Administrative Datasets in the LFE

- Business Register (BR)
- Longitudinal Employment Analysis Program (LEAP)
- General Index of Financial Information (GIFI)
- T1 business
- Statement of remuneration paid (T4)
- PD7 (Payroll Deduction Accounts)
- Research and Development in Canadian Industry (RDCI)
- Value of Foreign Direct Investment
- Canadian Direct Investment Abroad
- Trade in Commercial Services
- Trade by Exporter Characteristics
- Patents (Canadian Intellectual Property office)

Survey Datasets in the LFE

- Survey of Electronic Commerce and Technology
- Survey of Research and Development in Canadian Industry (RDCI)
- Survey of Innovation and Business Strategy
- Survey of Advanced Technology
- Survey of Commercialization of Innovation
- Survey of Intellectual Property Management
- Survey of Financing and Growth of SME's
- Survey of Digital Technology and Internet Use

Developmental datasets and other linkage environments

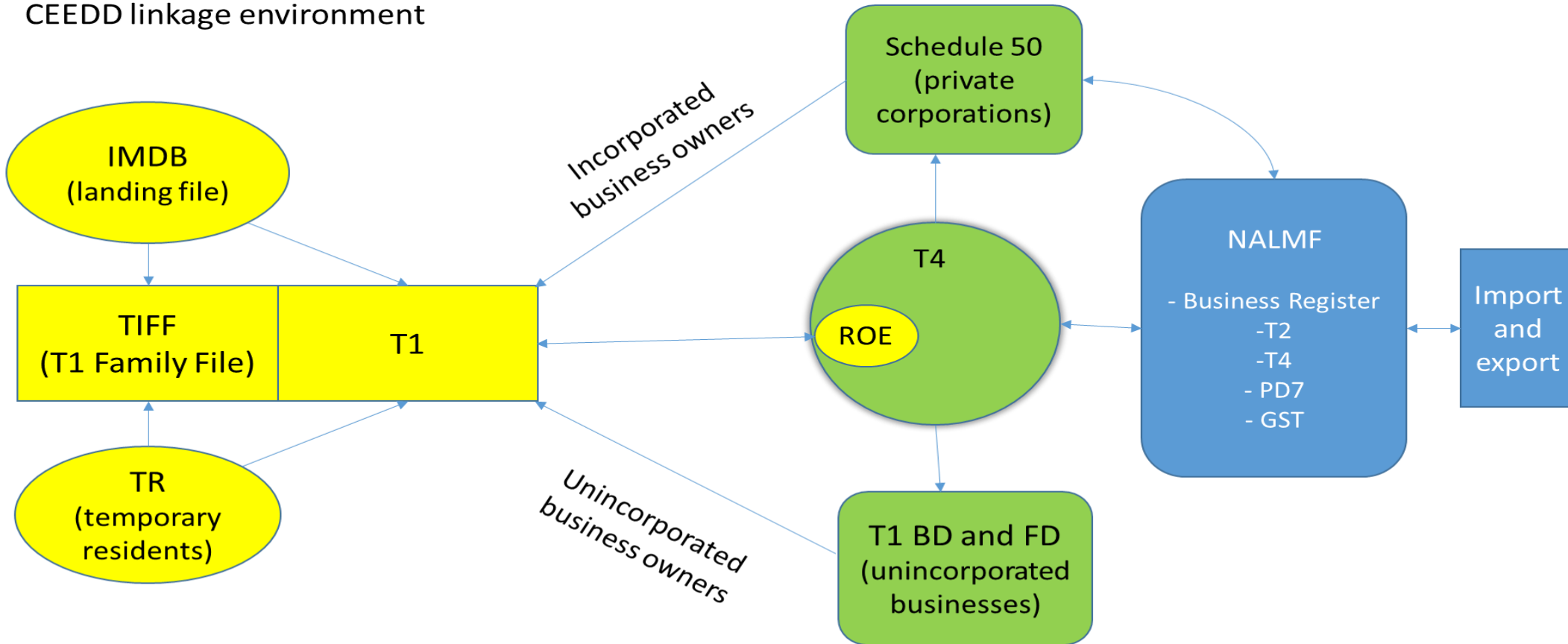
National Accounts Longitudinal Microdata File (NALMF)

- Longitudinal database of Canadian enterprises covering the 2000-2017 period
 - Successor to T2-LEAP file
 - Tracks a richer set of firm characteristics over time, such as employment, payroll, revenue, profit, assets, tangible capital stock, R&D capital stock, investment, value-added and productivity
 - Updated longitudinal structure; reconcile micro data with concepts and aggregates used and produced in the Macroeconomic Accounts
- Main data sources:
 - Statistics Canada's Business Register
 - Corporation Income Tax: T2
 - Employment: Payroll Account Deductions (PD7) and Statements of Remuneration Paid (T4 slip)
 - Goods and Services Tax: GST

Canadian Employer-Employee Dynamics Database (CEEDD)

- Linkage environment including links between employers and employees, from 2001 to 2016 (soon 2017)
- Covers the universe of tax filers, incorporated and unincorporated businesses in Canada
- Linkages of several tax administrative files:
 - T1 personal file, T1 family file, longitudinal immigration database, temporary residents file, record of employment
 - T4 statement of remuneration paid, T2 schedule 50, T1 unincorporated businesses
 - Data on firms productivity from NALMF, imports and exports

CEEDD linkage environment



Individual and family identifiers
 Business identifiers
 Individual and business identifiers

CEEDD: overview and use

- Contains information for 3 main fields:
 - **Employees** (demographic, immigrant status, employment income, etc.)
 - **Firms** (business type, employment, payroll, revenue, expense, profit, workforce, etc.)
 - **Business owners** (demographic, ownership type, ownership share, income from business)
- Possible Uses
 - Self-employment and business ownership
 - Interaction between workers and firms
 - Immigrant business ownership
 - Immigrants' initial firm allocation

Patents databases

- Canadian Intellectual Property Office (CIPO):
 - 1990 to 2012
 - Filing date, address, country, grant date, lapsed date, expired date, IPC classification
- US Patent and Trademark Office (USPTO):
 - 2000 to 2011
- Firm-level administrative data from NALMF
- Uses:
 - Patenting behavior – country of filing, joint filing, patent characteristics, firm characteristics
 - Innovation inputs and innovation outputs; innovation outputs and firm performance; IP use

Initiatives to improve access to business microdata

Current efforts to improve access to business microdata

Key motivation: bring **flexibility**, be client oriented

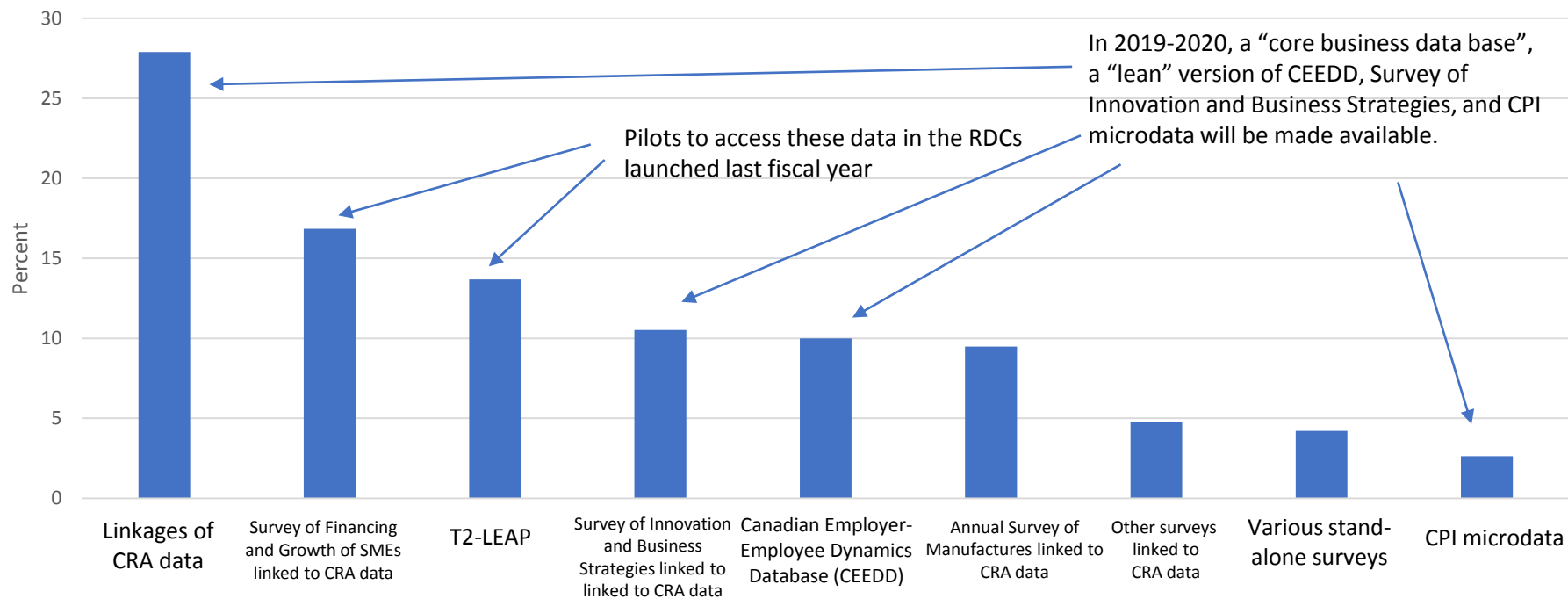
1. Business data pilot projects in the RDCs
2. Development of core business microdata

Business data pilot projects in the RDCs

- Pilot projects → business microdata into the RDCs under the current CDER model:
 - T2-Longitudinal Employment Analysis Program (T2-LEAP)
 - Survey of Financing and Growth of Small and Medium Enterprises (SFGSME) linked to tax data
- Why these two datasets?
 - They support a large fraction of the research proposals that have been made to CDER

Development of “core” databases

Distribution of data sources used in CDER



More information

- Website
 - <http://www.statcan.gc.ca/eng/cder/index>
 - <http://www.statcan.gc.ca/fra/cdre/index>

- Contact:
 - statcan.cder-cdre.statcan@canada.ca

Thank you!

Merci!