



Statistics Canada
www.statcan.gc.ca

Accessing Statistics Canada's Agri-Business Data for Research

Canadian Centre for Data Development
and Economic Research (CDER)

Brian Vockathaler
Economic Analysis Division

February 16, 2018

Outline

1. Canadian Centre for Data Development and Economic Research (CDER)
2. Datasets available through CDER
3. How to get access



Canadian Centre for Data Development and Economic Research (CDER)

- Repository for business and economic microdata files.
- Established in Ottawa in 2012 to provide secure access to external researchers to use business microdata at Statistics Canada for analytical projects.
- Application process required for access.
- Access is limited to the Statistics Canada headquarters in Ottawa.
- Cost recovery



CDER Agri-business datasets

1. Longitudinal Census of Agriculture (L-CEAG)
2. Survey of Innovation and Business Strategy (SIBS)
3. Trade by Exporter Characteristics (TEC)
4. Surface Transportation File (STF)
5. Longitudinal Annual Survey of Manufacturing (ASM)
6. Longitudinal Administrative Database (LAD)
7. National Accounts Longitudinal Microdata File (NALMF)
8. Canadian Employer-Employee Dynamics Database (CEEDD)



Longitudinal Census of Agriculture (L-CEAG)

- Longitudinal administrative database of farms and farm operators.
- Connects multiple censuses: 1986, 1991, 1996, 2001, 2006, 2011 and 2016 (in production)
- 1,531,706 records and currently has 101 variables.
- Longitudinally-consistent Agricultural Operation Identifier (AGOPID), industry- and geography-based classifying variables, and analytical variables.

Longitudinal identifier

- Agriculture Operation Identifier (AGOPID) is utilized as the longitudinal identifier between the individual censuses.
 - The AGOPID is part of each census-farm record
 - Identifies a farm, ranch, or other operation producing agricultural products for sale
 - A headquarters-based definition of the farm, rather than operator-based.
 - When a farm changes hands, but maintains the same address, the same AGOPID is maintained.

L-CEAG - Selected variables

Variable	Description	Variable	Description
<i>Matching</i>		<i>Inputs</i>	
AGOPIID	Agricultural operation identifier	FERTPD	Fertilizer and lime purchases
<i>Classification</i>		HERBCI	Use of herbicides
YEAR	Census year	INSECI	Use of insecticides and fungicides
PROV	Province	<i>Technology</i>	
LCSD	Longitudinal census subdivision	IRRIG	Use of irrigation
LNAICS	Longitudinal industry classification	COMPNY	Personal computers used for management
<i>Farm size</i>		TILLNO	No tillage
TFAREA	Total area of farms in acres	<i>Products</i>	
AOWNED	Area owned in acres	CANOLA	Canola
ALSDGOV	Area rented or leased from governments in acres	SUMMRF	Summerfallow
ARNTEED	Total area rented or leased from others in acres	TOTWHT	Total wheat
<i>Economic</i>		BARLEY	Total barley
SALE	Total gross farm receipts	TAMHAY	Total tame hay
TOTEXP	Total farm business operating expenses	TCATTL	Total cattle and calves
VALULB	Total land and buildings - market value	TOPIGS	Total pigs
VALMCH	Farm machinery and equipment - market value	CATLNY	Cattle on farm
TCSHWGE	Total wages and salaries	PIGSNY	Pigs



L-CEAG - Types of analyses

- Farm dynamics:
 - Trends in entry and exit over time.
 - Explaining farm entry or exit in Canada.
 - Life cycle patterns.

- Farm profitability:
 - The role of changing farm attributes over time.
 - By firm production function.

- Technology adoption:
 - The role of competitive pressures on adopting new seeding technology (tillage / no tillage).



Survey of Innovation and Business Strategy (SIBS)

- Cross-economy survey of businesses and industrial non-profit organizations in Canada. Two available vintages 2009 & 2012.
 - 2017 in production
- SIBS collects information on the strategic decisions, innovation activities, operational tactics and global value chain activities of businesses in Canada.

SIBS – Sampling

- The survey is limited to enterprises with at least 20 employees and revenues of at least \$250,000. Firms were sampled to meet the two following sets of objectives:
 1. Produce estimates that are representative by firm size (measured by employment) and industry not by geography.
 2. Permit microdata analysis within the 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.



SIBS – Analytical uses

- Firm / Enterprise level analysis on:
 - Obstacles to innovation
 - Innovation strategies – Cost cutting versus product diversification
 - Strategic focus of the enterprise
 - Advanced technology use

- Market analysis
 - Competitiveness

Trade by Exporter Characteristics (TEC)

- Replaces what is formerly known as the Export Register.
- Characteristics of Canadian businesses who export goods to countries outside of Canada.
- TEC has two data sources which are both linked to the Business Registry
 - Canada Border Service Agency (CBSA)
 - U.S. Census Bureau
- Enterprise level data, covering 2010 to 2016 (annual).
 - A firm must export a minimum of \$30,000 to be included in the dataset

TEC key variables 2010-2014

Enterprise

- Enterprise ID
- **NAICS**
- Province
- Census Metropolitan Area (CMA) codes
- Employment

Trade

- **HS-8 commodity codes**
- Country of destination
- Export only to US or not
- Number of partner countries
- Value of shipments

NAICS and HS 8-digit provide both industry and commodity dimensions.

Appropriate analysis and uses

- **Export dynamics**
 - Examining barriers to trade
 - Firms life cycle in the export market

- **Firm dynamics**
 - Propensity to export as function of R&D expenditure, owner demographics, firm size etc. (If linked to additional administrative data)

- **Industry or product concentration analysis**
 - Herfindahl Index

Surface Transportation File (STF)

- Built off the Trucking Commodity Origin and Destination Survey and Rail data
- Shipments characterised by **commodity**, tonnage, value and (network) distance
- Two major surface modes: trucking and rail
- Transportation costs measured on a level and ad valorem basis
- Period: 2004 to 2012
- Benchmarked to inter/intra-provincial trade totals
- Geocoded (lat/long) origin and destination points

STF features

- **Comprehensive.** Includes the majority of goods movement in Canada and between Canada and the U.S. by value
- **Flexible.** Inter-regional movement of goods on logistics *and* trade basis
- **Consistent.** Identifiers (i.e., geography and **commodity**) are consistent through time and across modes



Types of analyses

- Multi-use, micro-data file for analytical work
 - Cost of moving goods across the Canada-U.S. border
 - Provincial border effects
 - Canada-U.S. border effects
 - Market size available to firms
 - Effect of firm networks on trade flows
 - Inter-regional trade flows
- **SCTG/IOCC** codes can be used at an aggregate level.

Annual Survey of Manufacturing (ASM)

- Survey of manufacturing industries at the plant level (plant < establishment < enterprise)
- Quasi-Census
- Covers 1961 to 2012 *
- Consistent identifiers to follow plants and enterprises through time.



ASM Continued

- Variables include:
 - value of shipments, employment, salaries and wages, cost of materials and supplies used, inventories, commodity data, including shipments or consumption of particular products, exporter status... 100s of variables.
- NAICS codes can be used to focus on the agricultural sector, food manufacturing sector, food and beverage retailers
- Diversity of the file permits many kinds of analyses
 - Productivity
 - Firm dynamics
 - Exporting behaviour (interprovincial and international)



Longitudinal Administrative Databank (LAD)

- 20% sample of T1 Family File and Longitudinal Immigration Data Base, 1984 to 2015
- Census family and individuals
- Variables allow analysis on income and demographics, socioeconomic conditions.
- Has been used in the past to examine unincorporated farm-owners.
 - Self-employment identifier
 - Farm income



Why LAD?

- LAD is available in the RDC's
- Can be linked to Incorporated data (T2 – LAD)
 - Available at CDER
- Allows for different levels of aggregation of the family unit
 - Individual
 - Parent/Spouse
 - Family
 - Child
- Detailed geographical breakdown

National Accounts Longitudinal Microdata File (NALMF)

Longitudinal database of Canadian enterprises covering
the 2000-2015 period

- Tracks a rich set of firm characteristics over time, such as employment, payroll, revenue, profit, productivity, industry

- Main Data sources:
 - Statistics Canada's Business Register
 - Administrative data
 1. Corporation Income Tax: T2
 2. Employment: Payroll Account Deductions (PD7) and Statements of Remuneration Paid (T4 slip)
 3. Goods and Services Tax: GST



National Accounts Longitudinal Microdata File (NALMF)

- The database can be used to examine
 - Labour productivity
 - GDP and employment of firms across firm-size categories
 - Activity at sub-provincial levels (CA and CMA).

- Longitudinal Identifiers have been developed to study:
 - Entry and exits
 - Firms life cycle



Canadian Employer-Employee Dynamics Database (CEEDD)

- Matched employer-employee database, 1983 to 2012
 - More variables, especially at the firm-level, are available after 2001
- Covers the universe of individual tax filers, unincorporated businesses, and corporations in Canada.
- Links across various administrative tax files:
 - T1 individual tax, T1 Family File, Immigrant Landing File, T4 employment remuneration, T1 unincorporated business, Record of Employment, temporary residents file, firm performance data from the National Accounts Longitudinal Micro data File

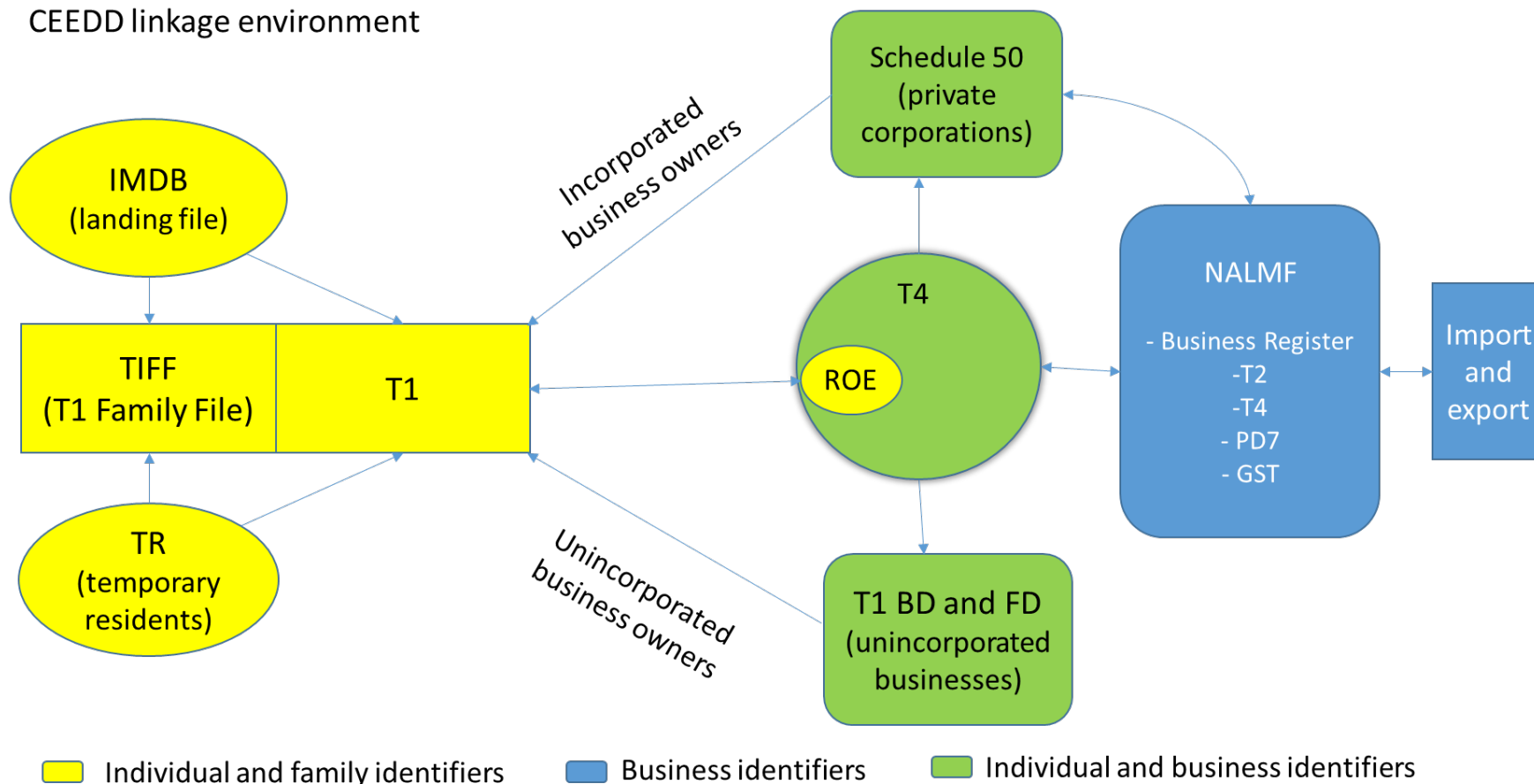


CEEDD continued

- Contains information for 3 main fields:
 - Employees (demographics, 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)
- **Can identify incorporated farms**
 - Interaction between workers and farms
 - Characteristics of the workforce

CEEDD – Linkage environment

CEEDD linkage environment





FAQ

1. Is it possible to link external data with Statistics Canada micro files?

- Yes it is possible link external files with Statistics Canada data. All record linkages request require approval form the Chief Statistician.

2. How much does it cost to gain access to CDER?

- All CDER projects begin with a base cost of \$ 10,000 and increase depending on the complexity of the project.



CDER access

- Three processes:
 - Researchers unaffiliated with a federal government department
 - Researchers affiliated with a federal government department
 - Researchers affiliated with a provincial government department

Non-federal government researchers

1. Draft a detailed proposal and contact CDER
2. Submit final project proposal
 - a. Includes application for accreditation
 - b. Letter from lead researcher indicating how costs will be covered, how peer review will be handled.
3. Evaluation of proposal
4. Security screening
5. Sworn in as deemed employee of Statistics Canada and sign a microdata research contract.



Contact information

statcan.cder-cdre.statcan@canada.ca



External Links

<https://www.statcan.gc.ca/eng/cder/index>

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