



Statistics
Canada

Statistique
Canada

Canada



Statistics Canada
www.statcan.gc.ca

Canadian Centre for Data Development and Economic Research (CDER)

Danny Leung, Economic Analysis Division
May 31, 2015

Outline

- What is CDER?
- Who is using it, what data have been used and what have researchers been working on?
- Application processes
- Procedures in place to ensure the confidentiality and privacy of the data
- Project costs
- Overview of some available data and ongoing data development projects



CDER

- CDER is the program that
 - provides approved analysts with secure access (in Ottawa, on a cost-recovery basis) to business microdata for research-oriented projects that serve the mandate of Statistics Canada
 - serves as a repository for, and spearheads the development of, confidential microdata involving businesses
 - provides analytical consulting services and support to users of business microdata.

Background (1)

- Statistics Canada has a mandate to produce statistics that help Canadians better understand their country.
- In the area of social data, Statistics Canada helped create the research data centres (RDCs) to address the need for professional assistance when carrying out work under the *Statistics Act*.
- For the same reason, CDER was created so that Statistics Canada could make better use of its business-data holdings without compromising security.

Background (2)

- In 2011, CDER was established at Statistics Canada's headquarters for federal government researchers and, in 2012, access was extended to non-federal government researchers.
 - As in the RDCs, an approval process is in place to ensure that projects meet professional standards, and researchers are qualified to undertake the work. Researchers become deemed employees of Statistics Canada to work on projects.
 - In addition, given the greater threat to the statistical system of a possible breach in confidentiality, CDER operates with measures in place to prevent data fishing.
 - Researchers work with synthetic data to write and validate their programs, and then submit their programs to be run in a monitored, batch-mode system.

Background (3)

- To facilitate the early commencement of projects, CDER was launched before data and documentation achieved the standard that exists in the social data deposited in the RDCs
 - Unlike social data, the production of business data has not been organized with the expectation that they would be widely used by individuals outside of Statistics Canada
 - Core data sets need to be identified and developed, as well as data documented and user guides written
 - Until then, researchers will need to work closely with Statistics Canada staff to use this data appropriately



CDER users

Government and government agencies

- Agriculture and Agri-Food Canada
- Atlantic Canada Opportunity Agency
- Bank of Canada
- BC Ministry of Finance
- Business Development Bank of Canada
- Citizenship and Immigration Canada
- Conseil Emploi Métropole – Emploi Québec
- Environment Canada
- Finance Canada
- Foreign Affairs, Trade and Development
- Industry Canada
- Parliamentary Budget Office
- Public Safety Canada
- National Research Council
- Natural Resources Canada

Other institutions

- Brock University
- CD Howe Institute
- Canadian Venture Capital Association
- Ivey Business School
- Kellogg School of Management
- Ryerson University
- Stern School of Business
- Telfer School of Management
- University of British Columbia
- University of California
- University of Calgary
- University of Guelph
- University of Ottawa
- University of Texas
- University of Washington



Some data sets at CDER (1)

- Annual Survey of Manufacturing
- Annual Survey of Manufacturing – Import Registry Database
- Annual Survey of Manufacturing – National Pollutant Release Inventory
- Bioproducts Production and Development Survey
- Canada Border Service Agency Customs Database
- Canadian Direct Investment Abroad
- Capital and Investment Program
- Canadian Employer Employee Dynamics Database
- Consumer Price Research Database
- Farm Environment Management Survey
- Financing and Growth of Small and Medium Enterprises
- Foreign Direct Investment in Canada
- Functional Foods and Natural Health Products Survey
- Industrial Product Price Microdata
- Longitudinal Agricultural Census
- Longitudinal Employment Analysis Program



Some data sets at CDER (2)

- Longitudinal Worker File
- National Accounts Longitudinal Microdata File
- Research and Development in Canadian Industries
- Survey of Advanced Technology
- Survey of Commercialisation and Innovation
- Survey of Digital Technology and Internet Use
- Survey of Drinking Water Plants
- Survey of Electronic Commerce and Technology
- Survey of Financing of Small and Medium Enterprises
- Survey of Innovation
- Survey of Innovation and Business Strategies
- Survey of IP Management
- T2-LEAP
- T2-LEAP-Import Registry Database
- Transportation Cost Database based on Trucking Commodity Origin Destination Survey and Rail Data
- Workplace Survey
- Workplace Employee Survey
- Workplace Employee Survey and Annual Survey of Manufacturers

Research themes (1)

- Firm financing
 - determinants of leverage for Canadian private and public firms
 - performance of venture capital in Canada
 - financing profiles of R&D performers
 - trends in financing activities among employer small and medium enterprises (SMEs)
 - heterogeneous credit conditions among R&D-performing SMEs
 - firm financial constraints and productivity
 - demand for finance, financial frictions, and trends in firm savings
 - economic impact of venture capital
 - impact of angel investment on company performance

Research themes (2)

- Innovation, firm strategies and firm performance
 - efficiency, allocation and competition
 - are innovative firms constrained by regulation
 - drivers of innovation, complementarity in innovation and performance
 - impact of management practices on performance
 - impact of innovation and advanced technology on profitability
 - determinants of SME business performance
 - foreign patent protection rights encourage domestic R&D spending?
 - value of intellectual property rights
 - open innovation improve firm performance?
 - heterogeneity in firms' R&D efforts in Canada
 - marketing innovations and the performance of businesses
 - importance of governance structures to firm performance



Research themes (3)

- Firm profiles
 - performance of immigrant-owned, newer firms
 - Canadian firm-size distribution, basic facts
 - high growth and rapidly shrinking firms
 - characteristics of Canada's top performing small businesses
 - Canada–U.S. productivity gap
 - Canada–U.S. capital stock comparison
 - growth processes of Canadian firms
 - growing profitable firms
 - risk-management perspective on multiple-business ownership
 - high-growth firms in Canadian industries
 - analyzing the dynamics of SMEs' growth transition
 - productivity in engineering services

Research themes (4)

■ International trade

- impact of international factors on Canadian retailers
- global value chains
- globally trading firms in Canada: Global value chains and productivity
- Canadian exporters—firm, country and product dynamics
- exporters and importers: a portrait of trading firms for Canadian goods
- internationalization facilitate or inhibit firm survival?
- theoretical and empirical examination of the exposure of Canadian retailers to cross-border shopping
- export density and internationalization, a contingency view on survival and growth of Canadian firms
- outward direct foreign investment (FDI) and firm performance
- multinationals, off-shoring, relocating and economic performance

Research themes (5)

- Evaluation of programs and policies
 - effectiveness of R&D direct grants versus R&D tax credits
 - economic impact of the Canada Small Business Financing (CSBF) Program
 - economic impact of the Industrial Regional Benefits Policy
 - cost benefit analysis and incrementality study of the CSBF Program
 - impact of corporate taxes on the growth of SMEs
 - National Research Council client impact measurement study
 - environmental regulation and competitiveness—evidence from the BC carbon tax
 - innovation induced by environmental regulation
 - assessing the growth barriers created by the enhanced R&D tax credit
 - effectiveness of trade commissioner's services
 - impact of the Canada–Chile free trade



Research themes (6)

- Prices
 - dynamics of the consumer price-distribution
 - price distribution and monetary non-neutrality in sticky price models
 - a quality bias in the CPI?
 - indirect trade and exchange rate pass-through
 - understanding incomplete exchange rate pass-through
 - durable goods pricing
 - effect of price level shocks on firm dynamics
 - import pricing and currency invoicing

Application processes

- There are two processes:
 - Researchers unaffiliated with a federal government department
 - Researchers affiliated with a federal government department
- Each process ensures that, if their project is accepted, the project falls under the mandate of Statistics Canada

Process for non-federal government researchers

Step 1: Draft a proposal and contact CDER

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.

Process for non-federal government researchers

Step 2: Submit project proposal

a) Application for accreditation:

- CV that demonstrates experience and technical competence;
- For graduate students, a letter of support from thesis supervisor;

b) A letter from lead researcher indicating:

- how project costs will be covered;
- how the peer review of the project will be handled;
- 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 an output for Statistics Canada.



Process for non-federal government researchers

Step 4: Complete the security screening process

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



Protecting confidentiality and privacy

- Safe projects
- Safe people
- Safe workplace
- Safe data
- Safe outputs



Safe people

- Researchers need to become deemed employees of Statistics Canada
- Researchers must be accredited
- Application must include how project costs will be covered and researchers have no conflict of interest to declare
- Orientation session on confidentiality



Safe workplace

- CDER is located at HQ on the 18th floor of RH Coats building
- Access to data on a dedicated server is through computers with disabled external media devices and with no links outside of Statistics Canada
- No printer privileges
- Project-person specific logins

Safe data

- Researcher databases are stripped of identifying information and scrambled ids are used
- Only variables necessary for the completion of the project are given to the researcher
- Two-stage access system
 - First stage – researchers are given a shuffled database to learn the structure of the data and to develop programs
 - Second stage – researchers use the actual data non-interactively, submitting their programs to run in batch mode

Safe output

- An analyst is assigned to follow each project
- A generalized tool developed at Statistics Canada is used to review outputs for confidentiality
- Results are shared with subject matter divisions to guard against residual disclosure
- A record of outputs are kept to help protect against residual disclosure

Project costs

- The cost of a project is determined on a project-by-project basis
- A simple project will cost approximately \$7,200
 - Project makes use of an existing database
 - 3 months (66 days) of access to CDER over a one-year period
 - It takes no more than 3 days for a Statistics Canada analyst to review the outputs of the project for confidentiality issues



Development of databases for analysis

- Longitudinal Employment Analysis Program (LEAP)
- T2-LEAP
- T2-LEAP-Import Registry
- Annual Survey of Manufacturers (ASM)
- ASM-Import Registries
- ASM-Workplace and Employee Survey (WES)
- National Accounts Longitudinal File
- Capital and Investment Program
- Longitudinal Worker File (LWF)
- Occasional special surveys (Survey of Innovation and Business Strategy, Survey on Financing and Growth of Small and Medium Enterprises)
- National Accounts (Productivity statistics, Input-Output Tables at a detailed industry and commodity level, Matched Canada-U.S. price data)
- Trucking Commodity Origin and Destination Survey
- Canadian Employer-Employee Database

LEAP and T2-LEAP

- Longitudinal Employment Analysis Program (LEAP), 1981 to 2012
 - Created out of T4 *Statement of Remuneration Paid* and data from the Survey of Payroll, Employment and Hours (SEPH).
 - Developed to study employment dynamics at the firm level. Source for CANSIM tables 527-0001 to 527-0012.
 - Dealt with changing firm structure by using labour tracking, a methodology that examines the movement of clusters of workers from firm to firm over time to track relationships across firms over time. Labour tracking was applied primarily to defining entry and exit, since this was the focus of the first studies.
- T2-LEAP, 1984 to 2012
 - Original LEAP files were expanded to include financial data to allow for a wider range of studies that required more than just employment change.
 - But the number of variables from the corporate tax system was initially limited because only 6 “tombstone” variables were available at Statistics Canada
 - Wealth of information that could be added expanded in the 2000s.

Annual Survey of Manufactures

- Cross-sectional, establishment level survey of manufacturing and logging industry.
 - Principal industrial statistics (revenue, employment, salaries and wages, cost of materials, cost of energy and wages, etc).
 - Take all, take some sample, administrative records for small units – sampling strategy by province, industry and revenue size has changed over time.
 - Although micro data back to 1961 exist, various longitudinal databases have been constructed with the help of industry and identifier concordances;
 - 1961-1990 (1970 SIC; 2-digit level); 1970-1990 (1970 SIC; 4-digit level)
 - 1961-1999 (1980 SIC; 2-digit level); 1973-1999 (1970 SIC; 4-digit level)
 - 1990-2010 (NAICS, standard ASM variables)
 - 2000-2010 (NAICS, UES ASML variables)



National Accounts Longitudinal Microdata File

- While the Input-Output Accounts provide estimates of GDP at the industry- or national-level, they are not available at a more disaggregated level, such as the sub-provincial level and the size of the firm.
- To estimate this, the National Accounts Longitudinal File (NALMF) combined a number of data sources, including administrative tax records (T2 and T4), data from the Business Register and from the Survey of Employment, Payrolls and Hours.
- It currently covers the period 2001 to 2012.
- Current longitudinal databases (i.e. LEAP) do not include detailed information on structural changes in the firm. The new database will include year-on-year details on structural changes.
- The data base can be used to track the GDP, employment, and financial statistics of firms and the locations where they operate. It can also be used to examine the size of firms in terms of GDP and employment, as well as their productivity.



Main data sources for NALMF

T4 Statement of Remuneration Paid

- Issued by employers, both incorporated and unincorporated, to employees for income tax purposes.
- Provides employment income, tax deductions, and province of employment.
- Contains both individual and business identifier. Primary source for LEAP.

T2 Corporate Income Tax

- Businesses include financial statements when they file taxes.
- Income statement provides data necessary to compute operating surplus for each corporate business.

PD7

- Businesses file employment income, deductions and employment with CRA (weekly, biweekly, monthly). Source for infra-annual data.
- Basis of employment numbers from the Survey, Employment, Payroll and Hours.



Main data sources for NALMF

Tax-filing entities identified by Business Numbers (BN) by the CRA may have operations across many geographical areas and in a number of industries.

The Business Register (BR) is needed to bring these administrative data to the level at which they can be compared with the SNA aggregates and other statistics published at Statistics Canada.

Business Register (BR)

- unduplicated list of all active businesses in Canada
- sampling frame for business surveys
- provides operating structure of each business
- provides the geography, revenue, employment and industry of all of the “operating entities” of each business



- Ongoing development of NALMF
 - Improve the methodology for creating a longitudinal file that allows the continuity of a firm across time and the relation of firms across time to be examined, and create a taxonomy of changes in the business population.
 - Impute data for micro-economic analysis, and include additional data sources to create a complete file of the business population.
 - Link the core database to other files (e.g., Export and Import Register, National Pollutant Release Inventory, GHG emissions data, Foreign Direct Investment Abroad, Canadian Direct Investment Abroad, Capital and Investment Program)



Capital and Investment Program (CIP)

- The measurement of capital stock and investment at the *firm level* is important to understand firms' productivity as capital is a primary input in production processes.
- The Capital and Investment Program (CIP) is a cross-sectional file that contains information on the stock of and investment in tangible and intangible capital assets for each incorporated employer business in Canada, for each year starting in 2000.
- For each business, annual estimates of total capital stock and investment are compiled from various schedules of T2 corporate income tax returns.
- The CIP file can be used to support research on firm-level economic and productivity analysis.



Longitudinal Worker File

- The Longitudinal Worker File (LWF) is an administrative database designed to provide information on employment dynamics in Canada.
- The LWF contains information on the demographic characteristics, including age, sex and marital status, of Canadian workers, on the jobs they hold, the firms in which they work and the earnings they receive.
- The LWF is constructed by integrating data from four sources: the T1 and T4 files of the Canada Revenue Agency, the Record of Employment (ROE) files of Human Resources and Social Development Canada, and the Longitudinal Employment Analysis Program of Statistics Canada.

Other databases

- Canadian employee–employer dynamics
 - In partnership with the Social Analysis and Modeling Division, and sponsored by Citizenship and Immigration Canada, a universe of matched, employer–employee database was created by linking NALMF to personal income tax data and the immigration database, which is comprised of (i) a file of employees with information about their workplaces (ii) a file of business owners—both incorporated and unincorporated—and the performance of their businesses.
 - The database is used to examine the role of immigrant-owned businesses in employment growth
- Multinationals database
 - In partnership with International Trade, Accounts and Statistics Division, a database that can be used to analyze the activities of multinationals is being built from NALMF, intercorporate ownership, foreign direct investment in Canada, Canadian direct investment abroad, and other administrative sources.

Other databases

- Transportation costs database
 - total and *ad valorem* transportation costs between destinations in Canada and the United States by truck and rail
 - allowing studies on Canada–U.S. trade, and interprovincial trade
- Longitudinal Census of Agriculture
 - linking the 1986 to 2011 censuses in longitudinal format
 - allowing studies on entry and exit, changes in farm populations, productivity in the agriculture sector
- National Pollutants Release Inventory (NPRI) linkages
 - NPRI linkage to business data, to data on employee separations due to illness, and to individual data on longevity and hospitalizations