



Statistics
Canada

Statistique
Canada

Quality of the Business Register (BR)

www.statcan.gc.ca



CANADA 150

Telling Canada's
story in numbers

**UNECE Group of Experts on Business
Registers**

September 27-29, 2017

Canada 



Presentation outline

- BR quality in context
 - What makes for a quality register?
 - Refresher on Statistics Canada's BR and its sources and processes
- Measurement and mitigation of frame errors
 - Indicators of under-coverage, over-coverage
 - Traditional and new methods of quality assurance and control
- Future challenges and opportunities for BR Quality



Statistics
Canada

Statistique
Canada

Canada

Business Register Quality in Context

The BR as a central statistical frame

- The BR is **central** to the **wide array** of economic statistics produced by Statistics Canada
 - Gross Domestic Product
 - Employment statistics (# of employees, wages, work hours)
 - Industrial sales and production data
 - Producer prices
 - Industrial data
 - Business demography and related topics (business financing, success/failure)
 - Many, many other examples
- A **myriad of programs** use and rely on the BR
 - Industrial survey programs
 - Economy-wide survey programs
 - Data integration projects
 - Research and Data Analysis Programs
- It is often used as a **sampling frame**, providing frame listings, and data for industry and size stratification
- Increasingly, it is also used for **allocation** using revenue and employment data and data integration projects

How do we define success in contributing to quality statistics?

Dimensions of Quality (Statistics Canada QA framework)

Relevance

- BR data content, concepts and classifications are a basis for meaningful statistics

Accuracy

- Listings of active businesses provide good coverage of target populations
- Businesses attributes needed for classification, stratification and allocation are reliable (i.e. industry codes, employee accounts, revenue values, etc)

Timeliness

- Point-in-time snapshots align with reference periods prescribed by statistical programs and BR is maintained evergreen

Punctuality

- Processing and delivery of frame data to statistical programs corresponds to their production schedules with the most up to date information

Coherence

- Concepts and classifications are standardized
- Statistical programs use and update the BR data according to standard procedures and processes

Interpretability

- Up-to-date metadata provided to explain statistics incorporate concepts, definitions and methods of the BR

Accessibility

- Business demographic data compiled from the BR are available to users inside and outside Statistics Canada



Basic pre-requisites for a quality statistical register

- **Good, timely data**
 - Administrative data; survey feedback information; info collected by profilers/analysts
- **A good team (or teams) with multi-disciplinary participation**
 - Profilers, economic statistics analysts, methodologists, national accountants, IT professionals
- **Collaborative tools**
- **Standardized concepts classifications**
- **Robust data processing and staff editing/verification procedures (and staff training)**
- **Reliable and sufficient IT infrastructure**
- **Good governance**



Statistics Canada's BR: Some background

- **6 million active enterprises**, all created and updated on the basis of **tax data**
 - Weekly data transfers from the tax agency
- **3.2 million** are **registered** with the tax agency's business taxation programs, managed through a common tax identifier – **the Business Number (BN)**
 - For BN-based enterprises, we use **corporate income tax returns, payroll deductions, and sales tax remittances** to update key frame variables
- **2.8 million** enterprises on the BR are **sole proprietors or partners** that have **no BN** (not required if no employees or no sales taxes to pay);
 - these are maintained using personal income tax returns
- **Quality assurance of enterprise-level data** focuses on the **optimal and efficient use of tax information**, which in general accurately portrays the revenues, # of employees, and other frame variables
 - Some challenges to overcome (e.g. industries)



Distribution of enterprises

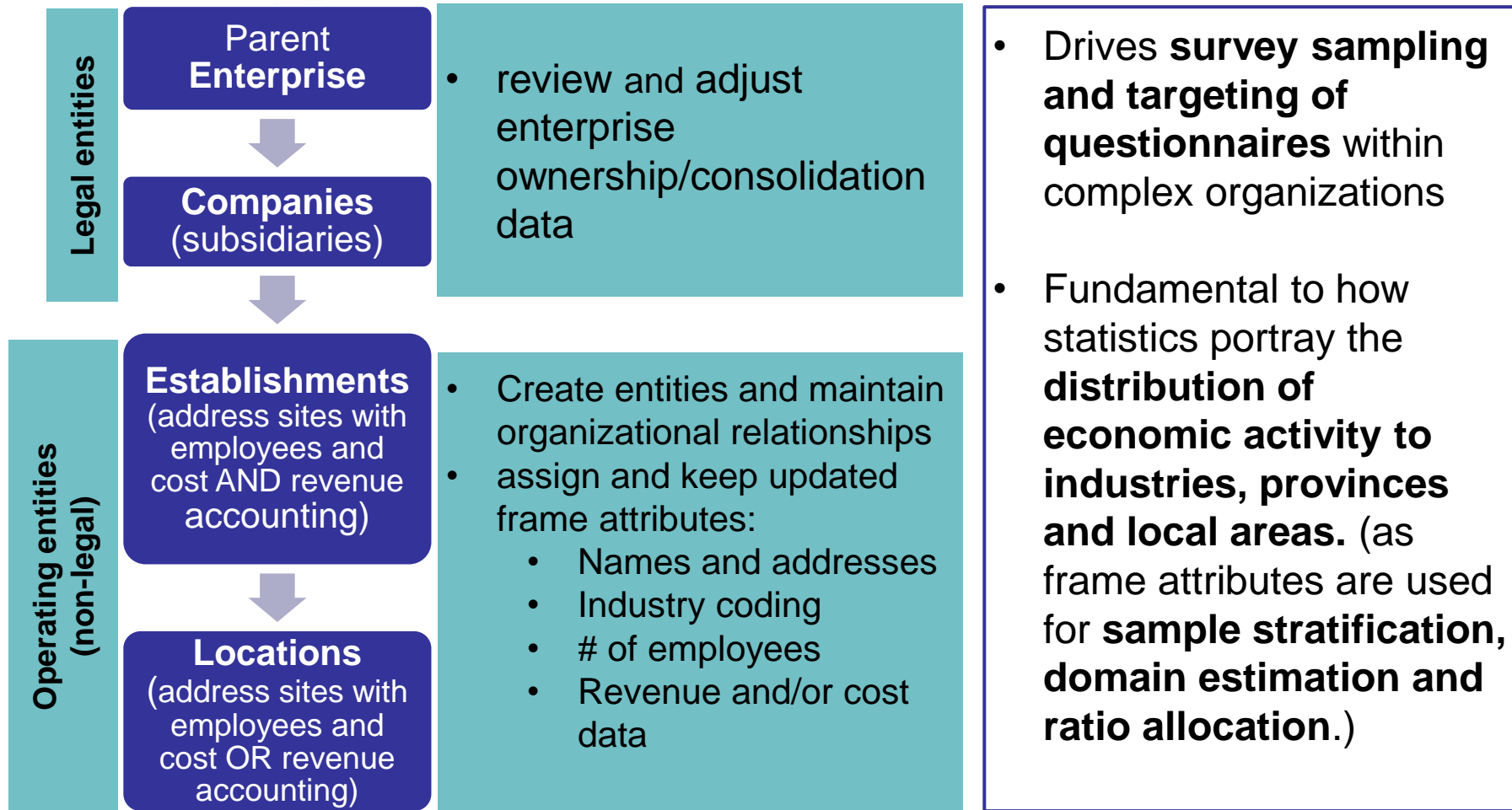
Business Type	Number of Enterprises	Generated Revenue	Employees
Simple	6 million (95%)	44%	56%
Profiled Large and Complex	22,000 (0.4%)	53%	41%
Administrative Complex	260,000 (4.6%)	3%	3%

22,000 profiled enterprises (53% of revenues)

Frame statistical units/structures...

...the work of profilers...

...and the need for quality





Statistics
Canada

Statistique
Canada

Canada

Measurement and mitigation of frame errors



How do we measure quality and maintain/improve accordingly?

- Various components work together
- 2-pronged strategy:
 1. Invest in independent and quality tracking/measurement activities, conscious of benefits vs costs
 2. Fully leverage existing data, taking advantage of opportunities to find anomalies through data confrontation/comparison

Frame under-coverage

- *Overall* under-coverage is not a major concern and not specifically measured
 - Registrations with the tax agency and use of remittances for income taxes, sales taxes and payroll deductions provides a reliable basis for identifying newly-active businesses
 - Business activities not reported for tax purposes are generally not part of the BR's target universe
- Under-coverage *within industries* can happen, due to:
 - **Missing industry codes** for some smaller businesses
 - **Misclassification** (measured by QA survey and other processes)
 - Complex profiles that do not fully and correctly elaborate the operating units engaged in different industries (specific quality control mechanisms)

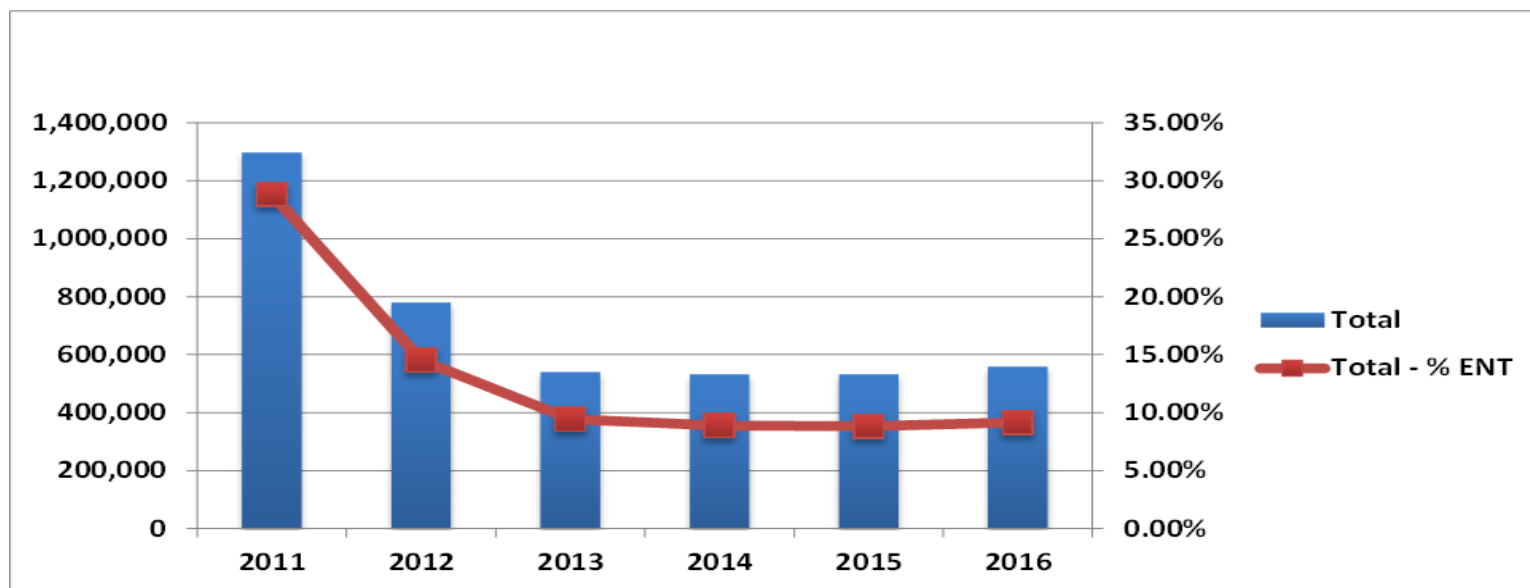
Frame over-coverage

- Can result from **time lags in the de-activation of units**, because:
 - Tax remittance accounts do not need to be specifically closed upon termination of business activities and income taxes can be reported in arrears
 - BR has wait periods once remittances have stopped prior to deactivating

- **Industry misclassification** causes under-coverage in one industry and over-coverage in another

Unclassified units (contributes to industry under-coverage)

Year	2011	2012	2013	2014	2015	2016
Counts	1,297,339	779,412	540,423	531,949	531,969	558,755
Ent. Population (%)	28.8%	14.5%	9.4%	8.9%	8.8%	9.2%





Results from BR Quality Assurance Program

- Annual sample of 3,600 businesses contacted to provide estimates of:
 - Missed de-activations
 - Misclassification

- For North American Industry Classification System (NAICS) *sectors* (20 groupings)

Industry misclassification

✓ Error

Average NAICS Misclassification Error Rate (ANMER)					
Year	2012	2013	2014	2015	2016
Based on Frequency	11.5%	12.4%	13.1%	14.7%	13.1%
Based on Revenue	0.0020%	0.0014%	0.0021%	0.0026%	0.0012%

✓ Volatility

Average NAICS Volatility Rate (ANVR)					
Year	2012	2013	2014	2015	2016
Based on Frequency	4.6%	3.3%	4.2%	4.0%	3.6%
Based on Revenue	0.0004%	0.0002%	0.0002%	0.0002%	0.0002%

✓ Total

Average NAICS Error Rate (ANER)					
Year	2012	2013	2014	2015	2016
Based on Frequency	16.1%	15.7%	17.3%	18.7%	16.7%
Based on Revenue	0.0024%	0.0016%	0.0023%	0.0028%	0.0014%

Industry misclassification

Average NAICS Error Rate (ANER) Based on Frequency and based on Revenue			ANVR	ANMER
Sector	Freq	Rev	Freq	Freq
11-Agriculture, Forestry, Fishing and Hunting	3.5%	0.0016%	1.1%	2.4%
21-Mining and Oil and Gas Extraction	23.0%	0.0012%	4.3%	18.7%
22-Utilities	15.7%	0.0001%	1.1%	14.6%
23-Construction	16.5%	0.0010%	1.5%	15.0%
31-33-Manufacturing	26.5%	0.0007%	5.6%	20.9%
41-Wolesale Trade	34.2%	0.0010%	10.2%	24.0%
44-45-Retail Trade	22.4%	0.0011%	6.9%	15.5%
48-49-Transportation and Warehousing	12.3%	0.0002%	3.7%	8.6%
51-Information and Cultural Industries	25.4%	0.0040%	4.3%	21.1%
52-Finance and Insurance	19.5%	0.0022%	1.0%	18.5%
53-Real Estate and Rental and Leasing	11.9%	0.0013%	3.4%	8.5%
54-Professional, Scientific and Technical Services	16.4%	0.0008%	4.2%	12.2%
55-Management of Companies and Enterprises	54.3%	0.0065%	5.1%	49.2%
56-Administrative and Support, Waste Management and Remediation Services	33.2%	0.0052%	5.9%	27.3%
61-Educational Services	17.0%	0.0030%	2.9%	14.1%
62-Health Care and Social Assistance	6.5%	0.0005%	1.6%	4.9%
71-Arts, Entertainment and Recreation	23.5%	0.0043%	4.9%	18.6%
72-Accommodation and Food Services	6.7%	0.0006%	2.1%	4.6%
81-Other Services (except Public Administration)	13.0%	0.0071%	4.0%	9%
Total	16.7%	0.0014%	3.6%	13.1%

New method for assessing NAICS codes

- QA survey sample size limitations
- We now have multiple sources for NAICS coding that can be used to 'confirm' one another:
 1. Tax agency registration
 2. Self-coding on annual income tax forms (for corporations and sole proprietors/partners)
 3. Verified/updated codes from 'survey feedback' (online questionnaire component)
 4. Updates by profilers and analysts

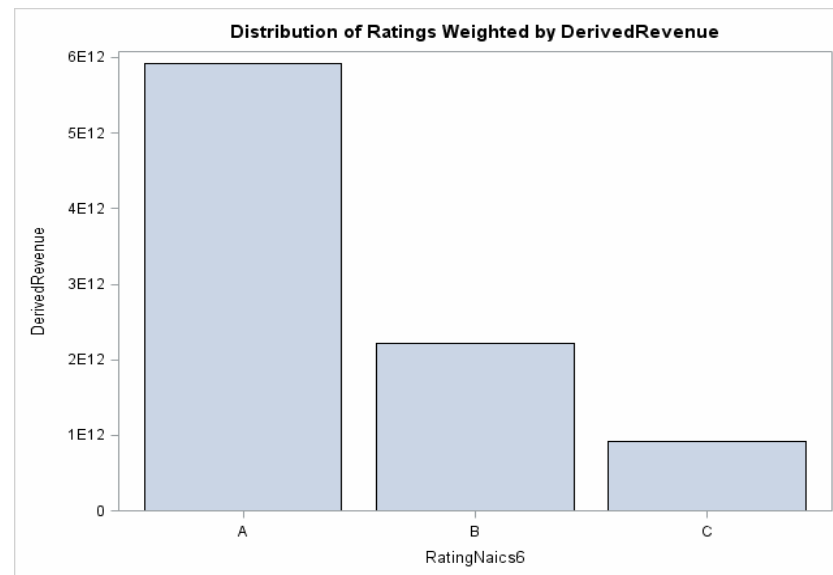
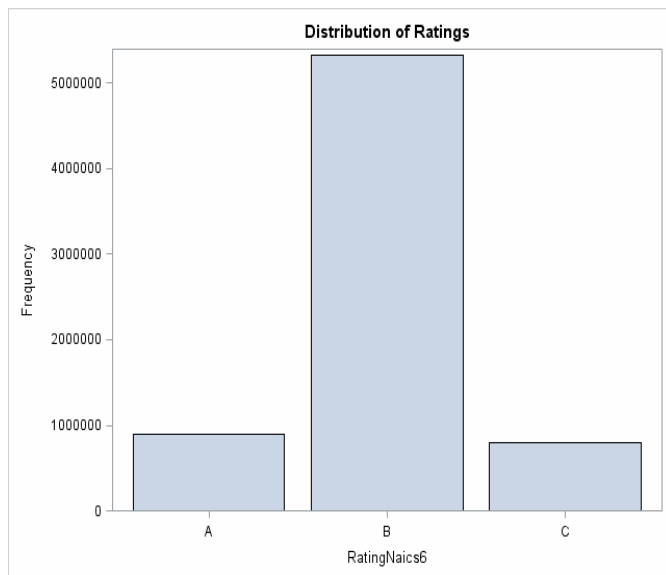
Composite rating of NAICS quality based on vintage of current value, number of confirmations and sector accuracy (QAP)

By counts

A	B	C
13 %	76%	11%

By revenue

A	B	C
65%	25%	10%



Frame duplication: on-going risks

1. (Unincorporated) sole proprietors sometimes have tax registrations for sales taxes or payroll remittances, but we don't always know the links
 - Can result in duplicate records on the BR; risks are attenuated for industries with higher concentration of small firms (e.g. agriculture)
 - Working intensively to improve record matching techniques, using better name parsing and introducing geo-locating methods
 - Also introducing collection of tax registration identifiers on survey questionnaires
2. Increasingly, businesses appear to be processing tax remittances through separately-registered tax accounts, sometimes through separately contracted firms
 - Profilers must take care not to double-count businesses, employees and revenues



Maintaining quality of profiler and analyst frame updates

- **Training** curriculum for updater/profiler; required for system privileges
- Close **collaboration** and consultation across **profiling teams and statistical subject-matter programs**
- **Quality Control module** recently implemented (audit recommendation) to sample and verify integrity of updates of staff from across Statistics Canada and follow-up as required

Other QA methods for complex profiles

- **Critical flag** prioritizes profiles for review (prior to December frame snapshot) and used in many annual surveys
- Each complex has a **composite integrity score**, based on:
 1. Additivity of (hierarchical) employment and revenue values (i.e. *vertical coherence*)
 2. Recency of last profile update
 3. Counts and types of 'messages' issued by analysts requesting reviews or other signals of possible changes
 4. Consistency of provincial employment distributions with annual summary of earnings and tax deductions sent to employees (T4 slips)
- Additionally, formalized methods are in place to compare BR frame variables (employment, revenues) with a variety of survey and administrative sources (*horizontal coherence*)

Current state in summary

- **Internal audit** of BR QA practices (2016) was **positive**
- **Frame coverage issues within industries are minor** in the context of economic activity; employees and revenues are being classified to the industries and provinces where they belong
- **Some issues** remain with “micro businesses” that have no identified links to the revenue agency’s business registrations and for which the precise moment of termination of business activity can be difficult to discern from the tax data
 - These issues will need to be mitigated as we build the longitudinal entrepreneurship database
- As the need **for local statistics** grows, we are increasingly challenged to find **more timely and efficient ways of elaborating complex enterprise statistical locations**



Statistics
Canada

Statistique
Canada

Canada

Future challenges and opportunities



The BR in a newly envisioned & modernized statistical system

- Newly-launched **Statistics Canada modernization** initiative challenges us to move to **admin/big data sources** wherever possible (instead of surveys) and ensure **collection and processing methods keep pace** with technology and advancements
- Heightens the role of **central registers as hubs for data linkage**; requiring **quality control for reconciliation of data records with register statistical units**
- We have begun working towards an **integrated system of core registers** of buildings, people and businesses; record linkage and frame quality assurance are key project components
- We are also looking for **alternative methods and sources for complex enterprise profiling**, such as outward facing web portal or usage of third party data; impacts on **data quality** will need to be taken into consideration