

PRASC



**Project for the Regional
Advancement of Statistics
in the Caribbean**

**Projet régional pour
l'avancement de la statistique
dans les Caraïbes**



In partnership with

Canada



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada

Statistical Business Register

Quality Assurance and Data Security
Best Practices

Project for the Regional Advancement of Statistics in the Caribbean (PRASC)

Jo-Ann Pedley and Gaétan St-Louis
November 1st, 2021, MS Teams



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada



Outline

- Quality Assurance Framework
- Canadian SBR Quality Indicators
- Data Security Measures
- Conclusion





Why a Quality Assurance Framework

1. Safeguard data quality
 - Timeliness, accuracy, coherence, relevance
2. Provide feedback to improve processes
 - Data processing procedures
 - SBR system interface



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada



Main Activities of Quality Assurance

1. Measure Quality
 - Coverage
 - Content (data/variables)
 - Maintenance processes and procedures
 - Systems and functionalities
2. Identification and analysis of quality issues
3. Prioritize changes required to improve quality
4. Communication – internally and externally

Know your data sets well



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada



Safeguards to Ensure Data Quality

Enforcing mandatory SBR training for all Statistics Office users

- Ensure a good understanding of the system and business rules
- Ensure to have common update procedures to keep the SBR data coherent
- Ensure a long-term sustainability for the SBR maintenance



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada



Safeguards to Ensure Data Quality

- Impact assessment on frame population variation
 - Month-to-month or year-to-year micro data comparison by using the Survey Universe File (SUF)
 - Month-to-month or year-to-year macro data comparison by using the demographic reports or producing custom tabulations from the SUF
- Identify any weaknesses from the update procedures
- Identify any potential problems with the administrative data as a source to maintain the SBR (e.g.: ISIC coding)
- Provide information to improve training material and update procedures





Safeguards to Ensure Data Quality

- Assess and measure the frame quality
 - Conduct a SBR Quality Assessment Survey to produce quality indicators
 - ISIC misclassification (ISIC error rate)
 - changes to ISIC (volatility rate)
 - Estimating the over-coverage (business no longer in operation)
 - Estimating the under-coverage
 - Unclassified units (missing ISIC)
 - Units with a prospect Flag=1 and not duplicate on the frame



Quality Indicators on Classification of Canadian SBR

Tables 2.2: Missclassification including EPMs and Previously Surveyed Units

Average NAICS Error Rate (ANER) incl. EPMs					
Year	2013	2014	2015	2016	2017*
Based on Frequency	15.7%	17.3%	18.7%	16.7%	12.9%
Based on Revenue	0.0016%	0.0023%	0.0028%	0.0014%	0.0013%

*2017 rate from June 2017 QAP (average of previous 12 months)

Average NAICS Volatility Rate (ANVR) incl. EPMs					
Year	2013	2014	2015	2016	2017*
Based on Frequency	3.3%	4.2%	4.0%	3.6%	3.1%
Based on Revenue	0.0002%	0.0002%	0.0002%	0.0002%	0.0001%

*2017 rate from June 2017 QAP (average of previous 12 months)

Average NAICS Misclassification Error Rate (ANMER)* incl. EPMs					
Year	2013	2014	2015	2016	2017**
Based on Frequency	12.4%	13.1%	14.7%	13.1%	9.8%
Based on Revenue	0.0014%	0.0021%	0.0026%	0.0012%	0.0012%

*Calculated using Average NAICS Error Rate minus Average NAICS Volatility Rate

**2017 rate from June 2017 QAP (average of previous 12 months)

Quality Indicators on Over-coverage of Canadian SBR

Table 2.1: Overcoverage including EPMs and Previously Surveyed Units

Average Death Rate (ADR) incl. EPMs					
Year	2013	2014	2015	2016	2017*
Based on Frequency	10.8%	11.3%	13.2%	11.8%	9.0%
Based on Revenue	0.0010%	0.0008%	0.0009%	0.0006%	0.0004%

**2017 rate from June 2017 QAP (average of previous 12 months)*



Quality Indicators on Under-coverage of Canadian SBR

Table 1.3: Undercoverage

Unclassified Units by Counts					
Year	2013	2014	2015	2016	2017
Total	N/A	507,920	504,883	525,515	538,619
Significant	N/A	212,400	222,454	229,403	247,130
% of Ent. Popn. Of Total	N/A	7.8%	7.7%	7.9%	8.0%
% of Ent. Popn. Of Significant	N/A	3.3%	3.4%	3.5%	3.7%

Unclassified Units by Derived Revenues					
Year	2013	2014	2015	2016	2017
Total	N/A	\$59,631,014,188	\$91,929,016,784	\$147,642,544,514	\$195,106,766,214
Significant	N/A	\$55,662,490,991	\$87,603,527,396	\$126,931,129,532	\$167,307,023,392
% of Ent. Popn. Of Total	N/A	1.1%	1.6%	2.5%	3.1%
% of Ent. Popn. Of Significant	N/A	1.1%	1.6%	2.1%	2.7%

Unclassified Units by Derived Employees					
Year	2013	2014	2015	2016	2017
Total	N/A	281,628	371,128	374,125	432,036
Significant	N/A	252,173	342,619	352,920	410,959
% of Ent. Popn. Of Total	N/A	1.5%	1.9%	1.9%	2.2%
% of Ent. Popn. Of Significant	N/A	1.3%	1.8%	1.8%	2.1%



Feedback to Improve Quality

- Feedback from QA could be used to improve:
 - Training and update procedures
 - BR system interface & update processes
 - Improve the quality of the administrative information



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada



Confidentiality and Security

- The SBR contains a lot of information on businesses received from various data sources
- The more data integration that is done, the more sensitive your micro linked database becomes
- Need to take additional security measures to protect your linked micro





Confidentiality and Security (Cont'd)

- It is recommended to put security measures in place to protect your linked data
 - Control the access to the statistical register (SBR)
 - Review the list of SBR users on regular basis (quarterly or yearly) to remove users who no longer require access
 - Store extracted data (SUF) in a protected directory on a file server with limited and controlled access (need to know basis)
 - Reduce to a minimum the number of copies of the extracted files (SUF)

Create a record linkage committee or a process to review, approve and document record linkage processes and requests for data





Confidentiality and Security (Cont'd)

- Think of de-identifying the extracted file (SUF) by removing direct business identifiers:
 - TaxId, SSId, RegistrarID etc.
 - Business name
 - Street address
- Develop statistical processes that will use the Statistical Number instead of the above direct business identifiers when possible





Conclusion

Factors to increase and maintain the quality of the SBR

- Good administrative data
- Good linkage processes
- Surveys using the BR as a frame
- Good training program

Challenges

- Up-to-date business profiling for large and complex businesses
- Maintain an accurate ISIC at the entity level
- Prospect flag review (to diminish the number of units with a prospect flag =1)

Secure your linked micro-database



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada

Statistical Business Register

SBR Maintenance Plan

Project for the Regional Advancement of Statistics in the Caribbean (PRASC)



Delivering insight through data for a better Canada

Jo-Ann Pedley
November 1, 2021, MS Teams



Statistics
Canada

Statistique
Canada

Canada



Objective

- Maintaining the SBR
- Creating a maintenance plan





Maintaining the SBR

- Important to maintain an unduplicated and up-to-date register
 - Know data sets well
 - Review data sets
 - Conduct analysis
 - Run clean-up processes: activation, inactivation, statistical flags
- Create a back-up
- Take a snapshot

Best practice = Create a Maintenance Plan

MONTH						
SUN	MON	TUE	WED	THU	FRI	SAT
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada



Maintenance Plan Best Practices

1

Create a schedule for the year

- Administrative data files
- Support activities
- Clean-up activities

2

Review and update schedule annually

- Changes in administrative data file delivery dates
- Pending NSO projects
- New activities

3

Create Back-ups and Snapshots

- File recovery
- Data loss prevention
- Survey Universe File (SUF)



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada



Create a Schedule for the Year

- Identify tasks that should be included
- Determine frequency of each task
- Store in an accessible area for users
- Update schedule to reflect the status of the task (completed, etc.)





Tasks to include in a Maintenance Plan

- Expected receipt dates of new administrative files (by partners)
- Dates administrative files should be loaded to the SBR
- Generating Survey Universe Files (snapshot)
- Creating system back-ups
- Inactivation and Reactivation
- Statistical Flag
- Others, as required



Example

August 2019						
Black = Scheduled <i>Blue/italic</i> = Executed Red = Cancelled Yellow = Tentative Server Maintenance						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
				1	2	3
4	5	6	7 <i>BR backup</i> <i>Administrative Data Update (NIS)</i>	8 <i>Administrative Data Update (NIS)</i>	9	10
11	12	13	14 National Hero's Day	15	16 <i>BR Backup</i>	17
18	19 <i>Server Maintenance</i>	20	21	22	23	24
25	26	27 Administrative Data Update (IRD)	28	29	30 BR Snapshot	31





Review and Update the Schedule Annually

- Update schedule each year
- Verify if changes to existing schedule are required
 - Date changes
 - Activity changes
 - New administrative sources



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada



Create Snapshots and Back-ups

Generate Survey Universe Files

- Survey sampling
- Longitudinal analysis
- Quality assurance
- Reporting

Best practice: At the end of each month

Back-up SBR

- Safety measure for file and data recovery
 - Hardware or software failure
 - Accidental deletion of file or data
 - Incorrect data modification
- Peace of mind

Best practice: Before an administrative load in case of loading issues



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada



You can contact the PRASC team at:
statcan.prasc-prasc.statcan@canada.ca

