

# PRASC



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

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

Funded by the  
Government  
of Canada

**Canada**

# Administrative Data Use

## Central Statistical Office of Trinidad and Tobago

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

Gaétan St-Louis  
Statistics Canada  
March 2020  
Port of Spain, Trinidad and Tobago



Delivering insight through data, for a better Canada



Statistics  
Canada Statistique  
Canada

Canada



## Outline

---

- Use of Administrative Data
- How to use Admin Files to Build a SBR
- Business Record Linkage
- Observations
- Recommendations
- Clarification Questions
- Conclusion

## What is a Statistical Business Register?

It is a repository of all businesses for a given country containing business tombstone information, size variables (Revenue, employment, etc.) and classification variables (ISIC, geography etc.) with other information

- An essential part of a country's economic statistics program and infrastructure
- A cost effective sample frame
- An important tool for stratification and estimation
- A central frame for integrating surveys and administrative data, thereby reducing response burden
- Finally, it is key in producing coherent GDP

# Use of Administrative Data

---



## Census versus Administrative Data

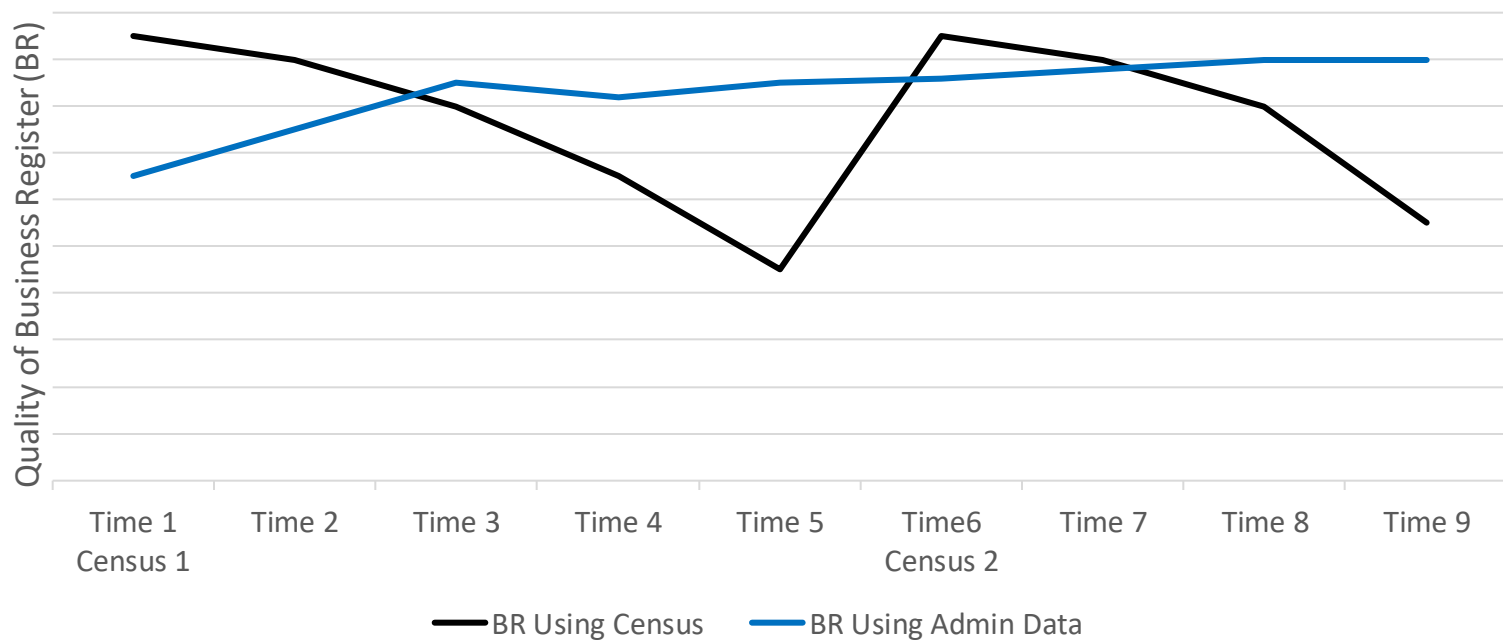
- Economic Census to create and maintain a Statistical Business Register (SBR)
  - Expensive and resource intensive
  - Reliant on funding
  - Burdensome
  - Subject to non-response (coverage)
  - Good coverage at census time
  - Quality issues over time (maintenance between two censuses?)
  - Good control of the content, definition and concepts
  - Difficult to assess longitudinal nature of the information

## Census versus Administrative Data

- Administrative Data to create and maintain a SBR
  - Inexpensive --> collected under an administrative program
  - Does not create additional response burden
  - Increases coverage depending on the sources
  - Stabilizes quality over time with continued scheduled maintenance
  - Easier to maintain longitudinal nature of the information (if unique identifiers available)
  - Limited or no control of the content, definition and concepts
  - Needs pre-processing of the data and record linkages

# SBR using Economic Census Data versus Administrative Data

Business Register data quality over time





## **Outcome of using Admin Data**

---

- Reduces response burden
- Lowers costs
- Improves quality of the SBR and
- Improves statistical programs

## Legal Framework

- The Central Statistical Office (CSO) has a **Statistics Act** that permits the access to documents or records required to accomplish their mandate, maintained in government departments or other organizations
- Administrative data received at the Statistics Office are protected under the **Statistics Act** in terms of security and confidentiality
- Statistics Act permits access to administrative data, however access is negotiated with admin partners and Memorandums of Understanding is usually prepared

## Memorandum Of Understanding

MOU includes :

- Acts of Parliament, if any, under which this sharing of information can take place
- Roles and responsibilities for the administration of the agreement
- Listing of data files obtained, proposed use by the agency and the frequency of the reception
- Predefined content and structure of data files to be shared
- Protection of information
  - provisions for maintaining **CONFIDENTIALITY** and **SECURITY** of the data
- Provision for audits

## Plan for Using Administrative Data

- Identify potential administrative data sources
- Review concepts, frequency and completeness
- Identify key variables of interest (tombstone, size variables, classification etc.) for:
  - Record linkage
  - SBR maintenance and
  - Survey data replacement
- Prepare a business case (justification for data use)
- Meet with administrative authorities to:
  - Determine files and variables that are available
  - Negotiate access

## **Plan for Using Administrative Data (Cont.)**

- Confirm the mode of transmission of data files (avoid Email)
- Prepare an agreement between NSO and each department (MOU)
- Develop a systematic quality assurance process
- Develop processes to align formats and concepts
- Develop record linkage processes (if no common identifier)
- Provide feedback to administrative authorities
- Schedule regular meetings with administrative authorities

## Tax Data Integration to the Canadian BR

### Canada Revenue Agency

Business Number  
Registration File

Corporate Tax Return File

Individual Tax Return File  
with Business Income

GST File

Payroll Deduction File

Secure  
transmission

### Statistics Canada

QA Process

Statistical Process  
(Edit, Calendarization)

Integrated  
Process

Statistical  
Business  
Register

## Key Business Register Variables on Canadian BR


Legal Base (TAX)	Statistics Canada's value-added activities		
<p><b><u>BN Registration File</u></b></p> <ul style="list-style-type: none"> <li>• Business Number (BN)</li> <li>• Legal/Operating names</li> <li>• Legal address</li> <li>• Telephone #</li> <li>• Contact name</li> <li>• Activity description</li> <li>• Legal type</li> </ul> <p><b><u>Other TAX files</u></b></p> <ul style="list-style-type: none"> <li>• Non-Profit indicators</li> <li>• International activity code</li> <li>• T2 Revenue/Asset</li> <li>• T1 Revenue</li> <li>• GST sales</li> <li>• Employment numbers</li> </ul>	<p><b><u>Profiling Activities</u></b></p> <ul style="list-style-type: none"> <li>• NAICS (industry)</li> <li>• Ownership links (subsidiaries)</li> <li>• Operating Entities and links</li> <li>• Operating Entities attributes (name, address, revenue, employment, NAICS, accounting center types etc.)</li> <li>• Ultimate Parent business</li> <li>• Ownership (foreign)</li> <li>• Country of Control</li> <li>• Consolidation information with attributes (NAICS &amp; sizes)</li> </ul>	<p><b><u>Generated</u></b></p> <ul style="list-style-type: none"> <li>• Unique Operating Entity number</li> <li>• Business status</li> <li>• Geography code</li> <li>• Allocation factors</li> </ul> <p><b><u>Statistical Indicators</u></b></p> <ul style="list-style-type: none"> <li>• Enterprise</li> <li>• Company</li> <li>• Establishment</li> <li>• Location</li> </ul>	<p><b><u>Collection Entity</u></b></p> <ul style="list-style-type: none"> <li>• Contact name</li> <li>• Address</li> <li>• Telephone number</li> <li>• Survey identifier</li> <li>• Frequency</li> <li>• Coverage</li> <li>• Response burden indicators</li> </ul> <p><b><u>Survey Sampling Specific Fields</u></b></p>

## Example on How to Use Admin Files to Build and Maintain a Business Register



## Example

---



Statistical Business Register (SBR)

## Example (cont.)

### Statistical Business Register (SBR)

BR Main Table								Link Table				
StatID	Bus. Name	Add.	ISIC	Rev	Emp#	Sales	Bus. Stat	StatID	REGISTID	TAXID	SSID	VATID
S0001	ABC	1 Main	5510					S0001	C1234			
S0002	CBC	2 Main	5610					S0002	C2356			
S0003	A&W	3 Main	5610					S0003	C4567			



- REGISTID
- Business Name
- Address
- ISIC

## Example (cont.)

### Statistical Business Register (SBR)

BR Main Table								Link Table				
StatID	Bus. Name	Add.	ISIC	Rev	Emp#	Sales	Bus. Stat	StatID	REGISTID	TAXID	SSID	VATID
S0001	ABC	1 Main	5510		10		Alive	S0001	C1234		N4210	
S0002	CBC	2 Main	5610					S0002	C2356			
S0003	A&W	3 Main	5610		20		Alive	S0003	C4567		N9365	
S0004	XYZ	5 Main	1030		65		Alive	S0004			N1367	



- REGISTID
- Business Name
- Address
- ISIC



- SSID (NIBID)
- Business Name
- Address
- ISIC
- # Employees

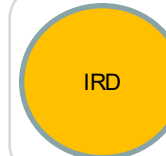
## Example (Cont.)

### Statistical Business Register (SBR)

BR Main Table								Link Table				
StatID	Bus. Name	Add.	ISIC	Rev	Emp#	Sales	Bus. Stat	StatID	REGISTID	TAXID	SSID	VATID
S0001	ABC	1 Main	5510	10000	10		Alive	S0001	C1234	R2001	N4210	
S0002	CBC	2 Main	5610					S0002	C2356			
S0003	A&W	3 Main	5610	90000	20		Alive	S0003	C4567	R2004	N9365	
S0004	XYZ	5 Main	1030	22600	65		Alive	S0004		R2005	N1367	



- REGISTID
- Business Name
- Address
- ISIC



- TAXID
- Business Name
- Address
- ISIC
- Revenue



- SSID (NIBID)
- Business Name
- Address
- ISIC
- # Employees

## Example (Cont.)

### Statistical Business Register (SBR)

BR Main Table								Link Table				
StatID	Bus. Name	Add.	ISIC	Rev	Emp#	Sales	Bus. Stat	StatID	REGISTID	TAXID	SSID	VATID
S0001	ABC	1 Main	5510	10000	10	14000	Alive	S0001	C1234	R2001	N4210	V3555
S0002	CBC	2 Main	5610			55555	Alive	S0002	C2356			V3666
S0003	A&W	3 Main	5610	90000	20		Alive	S0003	C4567	R2004	N9365	
S0004	XYZ	5 Main	1030	22600	65	38888	Alive	S0004		R2005	N1367	V3888
S0005	ZZZ	6 Main	3555			150000	Alive	S0005				V3999

**Registrar**

- REGISTID
- Business Name
- Address
- ISIC

**IRD**

- TAXID (BIRID)
- Business Name
- Address
- ISIC
- Revenue

**NIB (Employer Data)**

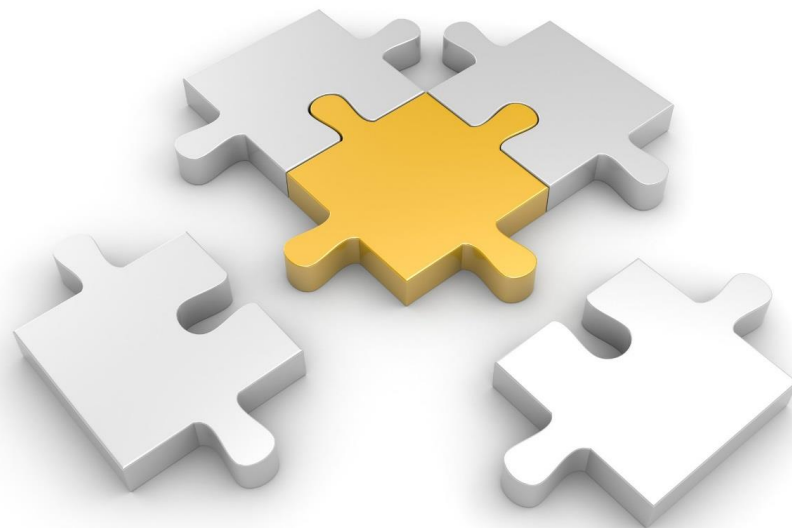
- SSID (NIBID)
- Business Name
- Address
- ISIC
- # Employees

**VAT**

- VATID
- Business Name
- Address
- ISIC
- Sales

# Introduction to Business Record Linkage

---



## Business Record Linkage (Cont.)

- BRL is a process of matching files together to:
  - Identify records in each file that represent the same person or business
  - Find duplicates in a single file by matching it to itself
  - Produce new datasets for analysis and maintaining the Statistical Business Register (SBR)

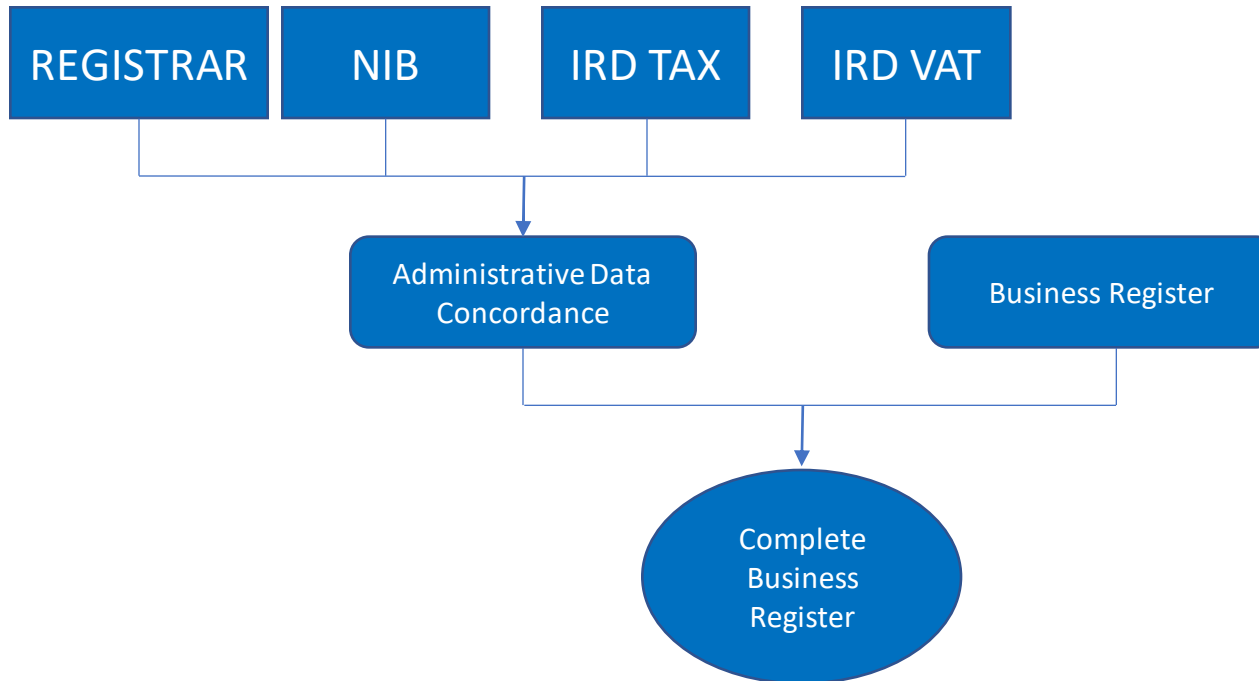


## Business Record Linkage (Cont.)

- Record Linkage Plan
  - Identify a base file for initializing the SBR
  - Identify the files you want to link to your base file
  - Identify linkage variables
    - Unique common identifiers (if available)
    - Business Names
    - Addresses
    - Telephone Numbers
    - Etc.
  - Identify a software (SAS, Excel Fuzzy Look-up Table etc.) and a methodology (deterministic or probalistic) to perform the record linkage

# Business Record Linkage

## Record Linkage Plan

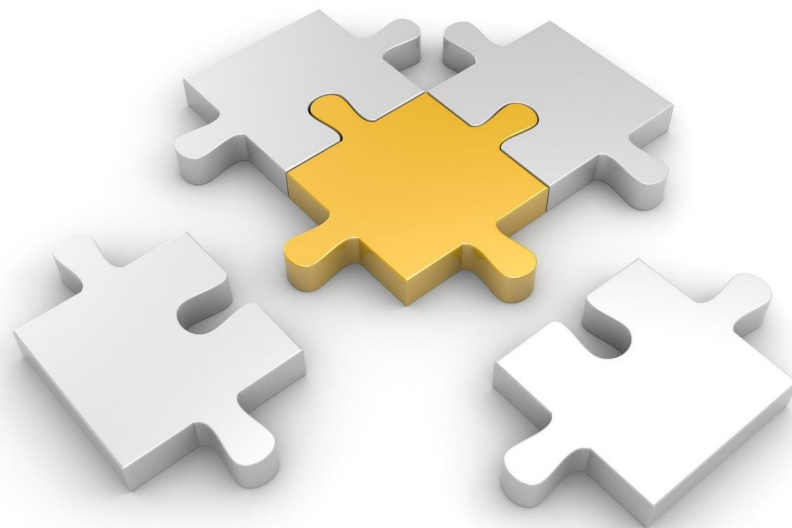


## Ongoing BR Update Process using Administrative Data

- Transform data to meet statistical needs (concepts & format)
- Develop a linkage process using linkable variables (business name, address etc.)
- Create a concordance table containing all administrative business IDs and unique statistical business IDs
- Identify key variables to be updated or created on the BR
  - Tombstone (business name, address, tel. # etc.)
  - Size variables (revenue, employees, salary and wages etc.)
  - Classification variables (geography, industry, others)
- Derivation of key variables
  - Develop business status rules (prospect, active, inactive, dissolved)
  - Create an “update date” and “source” variables for maintenance purposes

# File Preparation

---



## File Preparation

- Review concepts and completeness
  - Understand code sets
  - Understand definitions of variables
  - Identify key variables to use
    - For business record linkage
    - For SBR content and maintenance
  - Understand file contents
    - All requested variables present?
      - Are they well populated?
    - All key records present? (Large enterprises, very well known enterprises etc.)

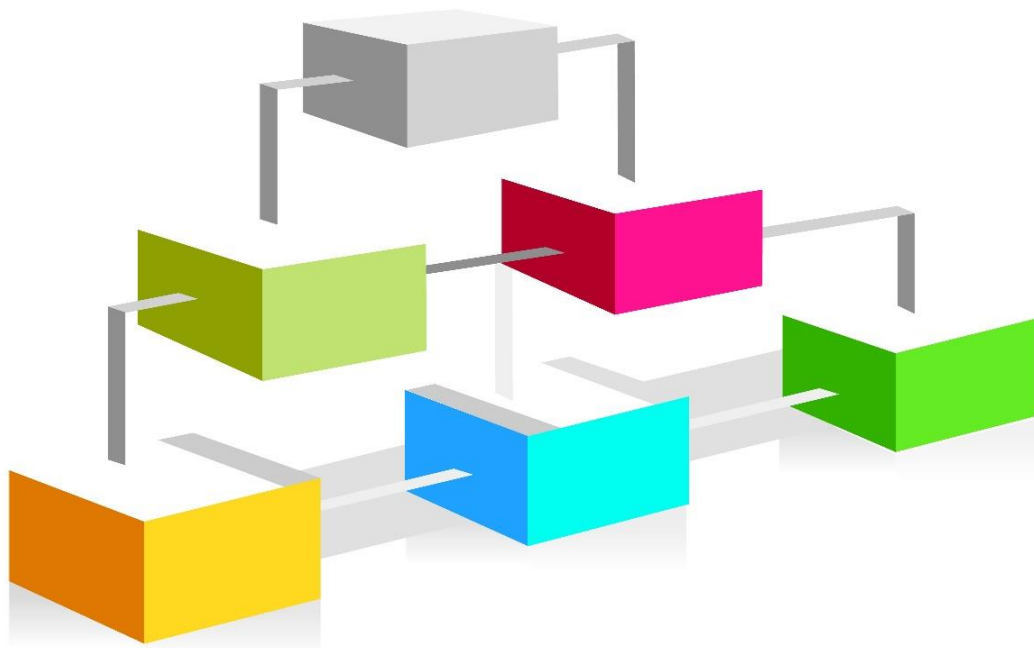
## Base File Identification

- How to identify a base file for initializing the SBR?
  - Good coverage of the business population
  - Well populated with at least a size variable (revenue or sale or number of employees)
  - Good tombstone information (Legal Name, Operating Name, Address, tel #, etc)
  - Good classification variables (ISIC and others)
  - Contains recent information
  - Generated on regular basis

## Annualization of Admin File

- Annualizing is the conversion of monthly or quarterly data (Sales or # of employees) to an annual value
- For VAT, the sales should be annualized
  - For a monthly remitter by adding the latest 12 months
  - For a quarterly remitter by adding the latest 4 quarters
  - For a yearly remitter, no action required
- For number of employees from the Social Security:
  - It is recommended to calculate an average number of employees from the latest 12 months of remittances
    - To reduce the impact of seasonality
    - To reduce the impact of error reporting
    - To lower the impact of monthly reporting with zero employee

## Observations and Recommendations



## General Observations

- No common unique identifier between administrative data sources (REGISTRAR, NIB, IRD Tax/VAT)
- Addresses are not standardized and post offices are often used
- Business names are not always reported consistently and sometimes contain:
  - Typography errors
  - Differences between name fields
  - Different record length (truncated fields)

## Recommendations

---

- To use a unique business identifier
  - If not possible, try to reuse existing registration numbers such as the TaxId or NIBID
- To standardize the addresses
  - Civic number (PO Box), Street Name, Street Type and City in separate fields
- To collect business name identifiers in separate fields
  - Example: Registered Name = 1234 Company Inc.  
Trade Name = Hotel Gaetan

## Advantages of Improving the Quality of Business Identifiers

- Facilitate management of administrative programs
  - Avoid the risk of having duplicate information
  - Ease the production of reports or monitoring
- Reduce response burden and cost
- Ease data integration in producing a reliable statistical infrastructure and high quality official statistics
- Facilitate continuous feedback between administrative departments on data quality improvements
- Improve record linkage results

# Questions

---



## Questions (1)

---

- Do any departments collect other program identification numbers (e.g. TAX to collect NIBID or REGISTRAR number)?
  - If not, would you be willing to collect it?
- Is all administrative data available in electronic format?
  - What are the variables captured?
  - When and how can you share these data sets with the Central Statistical Office (CSO)?

## Questions (2)

---

- Can a business have multiple branches/establishments and is this information registered and collected?
  - If so, in what scenarios?
- Does a sole proprietor need to be registered with your department?
- Is the business type (Sole Proprietor, Corporation, Partnership, etc.) available on the REGISTRAR, NIB, IRD Tax/VAT files etc.?

## Questions (3)

---

- Is the business status of a unit maintained by the administrative departments?
  - If so, how is it maintained and are there various statuses collected?
  - Is there a waiting period?
- Is there a way to link the legal enterprise to its establishments through the REGISTRAR, VAT, TAX or NIB files or other files?
  - If yes, are there other variables or indicators CSO should be requesting?

## Questions (4)

---

- Is the ISIC code collected and used?
  - Which version of ISIC and is the activity description captured?
- Will the CSO and admin partners be interested in collaborating in developing a common training package on ISIC coding?

## Questions (5)

---

- Does your office normally develop MOU when sharing data? Should a common standard template be developed?
- Does the admin partners have a secure mode of confidential data transmission. If not, what should be the mode of transmission?
- Is your office interested in developing a common business identifier to facilitate the administration of your program and production of official statistics using administrative data?
- What are the non-legislative issues your office may identify when sharing admin data with CSO and what will be your recommendations to reduce that barrier?

## Conclusion

---

- Use of administrative data by the CSO will help produce more reliable official statistics while reducing response burden and costs
- Collaboration between CSO and data providers will result in improving the quality of the administrative data (e.g. Industry classification)
- Continued work and communication with administrative data providers to identify potential new administrative data sources to be used by the CSO to improve their statistical programs and to identify potential new collaboration initiatives



You can contact the PRASC team at:

[statcan.prasc-prasc.statcan@canada.ca](mailto:statcan.prasc-prasc.statcan@canada.ca)

Canada 