

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



Administrative Data Use

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

Gaétan St-Louis

Date



Delivering insight through data for a better Canada



Statistics
Canada

Statistique
Canada

Canada



Outline

- Use of Administrative Data
- Integrating Administrative Data
 - Canadian experience
 - CARICOM possibility
- How to use Admin Data to Build a SBR
- Data Security and Confidentiality
- Observations
- Recommendations
- Conclusion

Use of Administrative Data





What is Administrative Data from a National Statistical Perspective

- It is all microdata collected outside of the National Statistical Office (NSO) such as:
 - Tax Department
 - Social Security Department
 - Central Register Department
 - Education Department
 - Health Department
 - Telecommunication Department
 - Etc.



Advantages of using Admin Data

- Reuses data already available:
 - Reduces response burden
 - Lowers collection costs → better use of public funds
 - Access to more data
- Improves statistical programs
 - More regular SBR updates
 - Potential for better survey estimates (more data available)
 - Potential for quicker turn around of estimates
- Improves the National Statistical System
 - More coherent official statistics in a cost effective manner
 - Members of the NSS working together to improve data quality



Micro Admin Data vs Aggregate Data (Cont'd)

Aggregate data:

- Useful as input to the production of National Account estimates
- Useful to produce some basic statistics

However, it greatly limits

- Developing a strong statistical infrastructure
- Capacity of statistical analysis, and
- Efficiency in producing official statistics



Micro Admin Data vs Aggregate Data (Cont.)

Micro Admin Data

- Permits statisticians to provide useful feedback to the admin partners on data quality and coherency
- Provides the possibility of applying internationally recognized statistical processes in order to make it fit for use for statistics such as: Quality Assurance Processes, Statistical Edits and Imputation, Annualization, Calibration Etc.
- Allows record linkage processes to develop new products, new official statistics, replacement of survey data and increase the capacity for statistical analysis



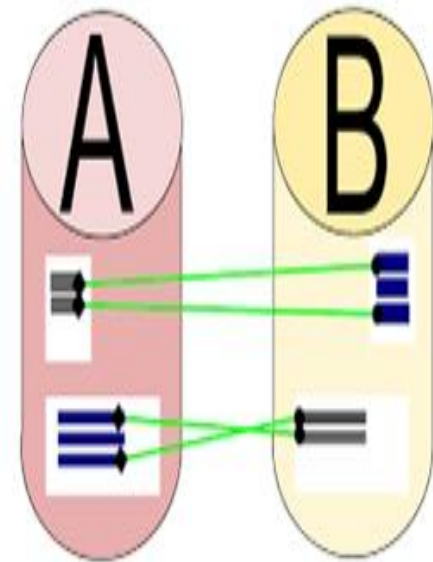
Micro Admin Data vs Aggregate Data (Cont.)

Micro Admin Data

- To rapidly produce Official Statistics for day-to-day decision making during times of pandemic, hurricane etc.
- Gives the capability of developing and maintaining a strong statistical infrastructure such as a SBR
- ***Ideally micro data are shared in electronic format with a secure transmission schedule with a signed MOU***

Record Linkage

- The process of matching files together to Identify records in each file that represent the same person, business, object or event
- Essential when using micro administrative data from various sources
- Issues that are often encountered:
 - Lack of unique common identifiers
 - Poor quality of unique common identifiers
 - Lack of standardized formats
 - Contain typographical errors
 - Large volume of records





Census versus Administrative Data

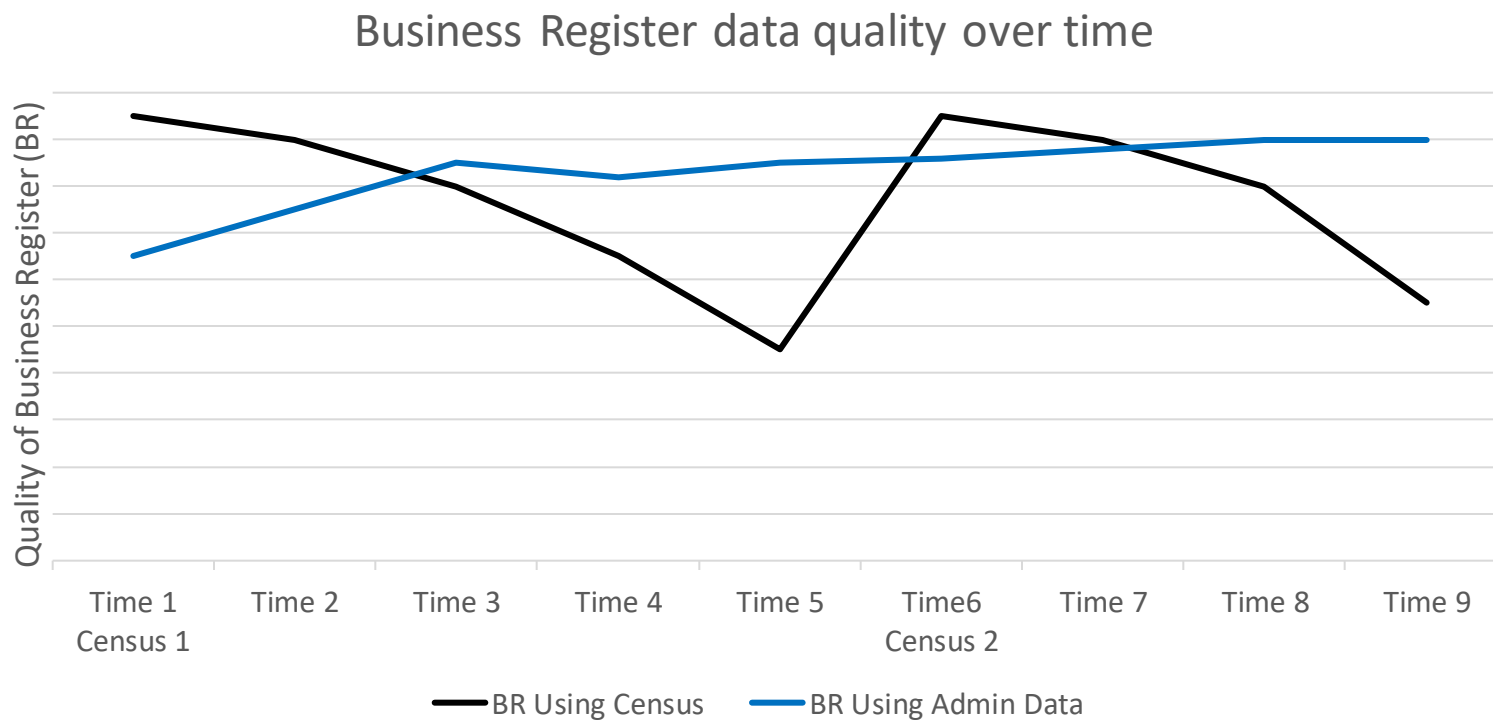
- **Economic Census** to create and maintain a Statistical Business Register (SBR)
 - Expensive and resource intensive
 - Reliant on funding
 - Burdensome
 - Subject to non-response (coverage)
 - Good coverage at census time
 - Quality issues over time (maintenance between two censuses?)
 - Difficult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



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 and record linkage
 - Stabilizes quality over time with continued scheduled maintenance
 - Easier to maintain longitudinal nature of the information (if unique identifiers are available)
 - Limited or no control of the content, definition and concepts
 - Needs pre-processing of the data and record linkages
 - Alignment with international practices

SBR using Economic Census Data versus Administrative Data





Legal Framework

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



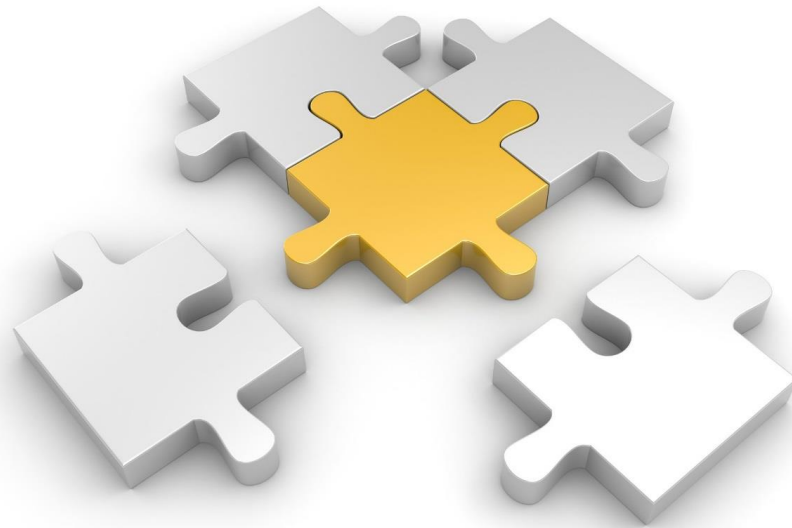
Memorandum Of Understanding

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

Try to move away from gentlemen's agreements

The Canadian experience with admin data to build and maintain a SBR





How is the SBR built in Canada?

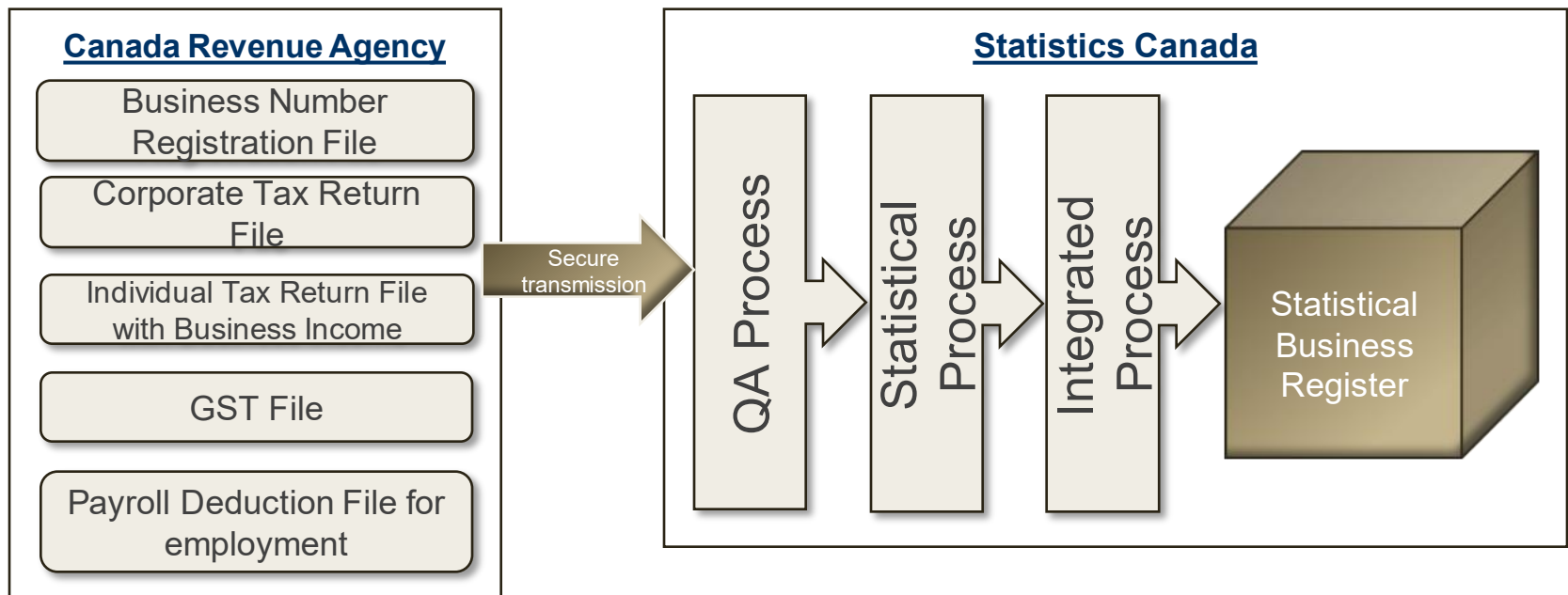
Administrative base using Canada Revenue Agency (CRA) information

- Business Number (BN) Registration File, complemented by
 - Payroll Deduction filers (PD7) for Employment Numbers
 - Corporate Tax Filers (T2) for Revenue and Assets
 - Individual Tax Filers Reporting Business Income (T1)
 - Goods and Services Tax Registrants (GST) for Sales
 - Other Tax Filers such as Partnerships and Charities

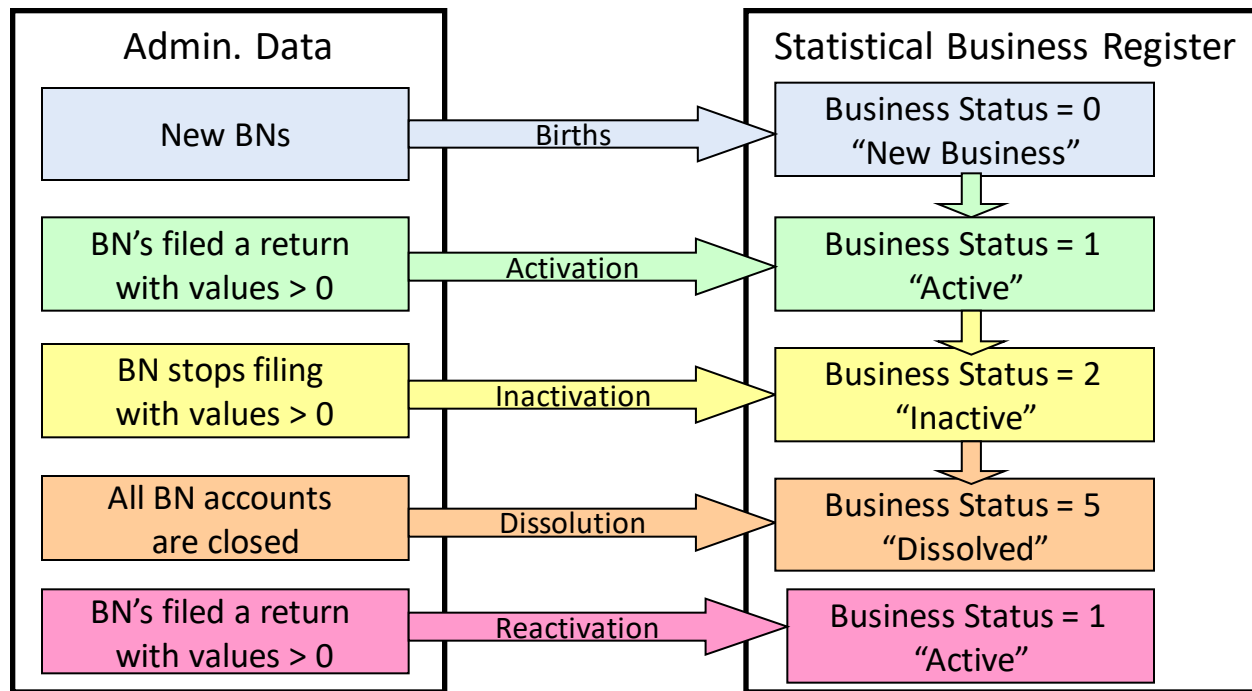
Value added activities

- Profiling activities - building Legal/Operating business structures
- Industrial (NAICS) and other classifications
- Maintaining a business status

Administrative Data Integration to the SBR in Canada

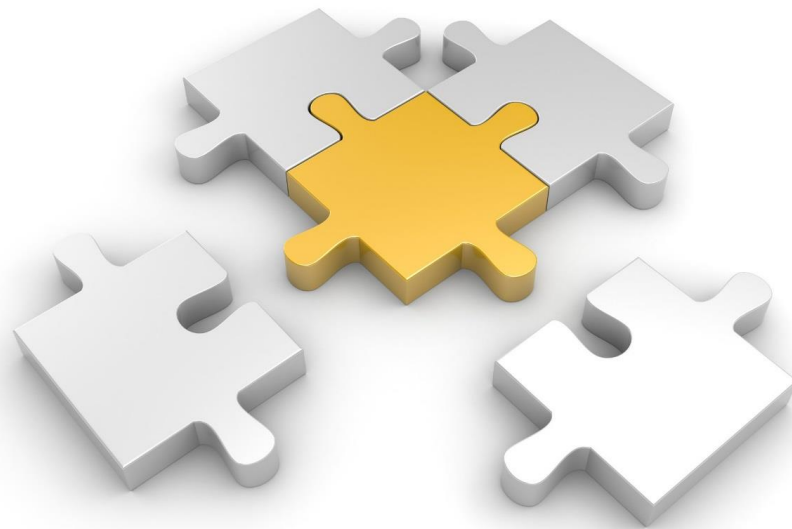


Business Status Flow in Canada





How can admin data be used to build and maintain a SBR in the CARICOM countries?





How can admin data be used to build and maintain a SBR in the CARICOM countries?

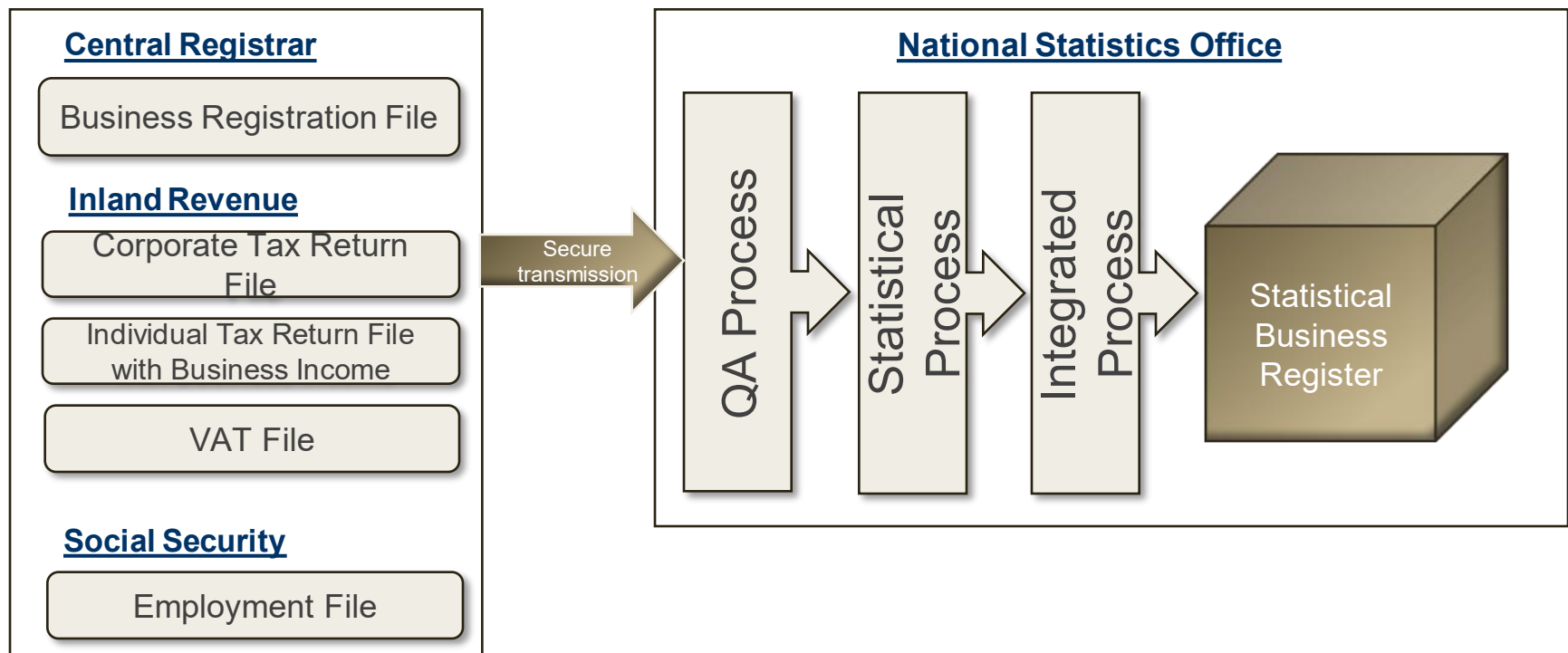
Administrative base using Inland Revenue, Social Security and Registrar information

- Social Security (SSD) for Employment Numbers & ISIC
- Inland Revenue (IRD) for Revenue and Assets & ISIC
- Value Added Tax (VAT) for Sales and ISIC
- Registrar for tombstone (legal)

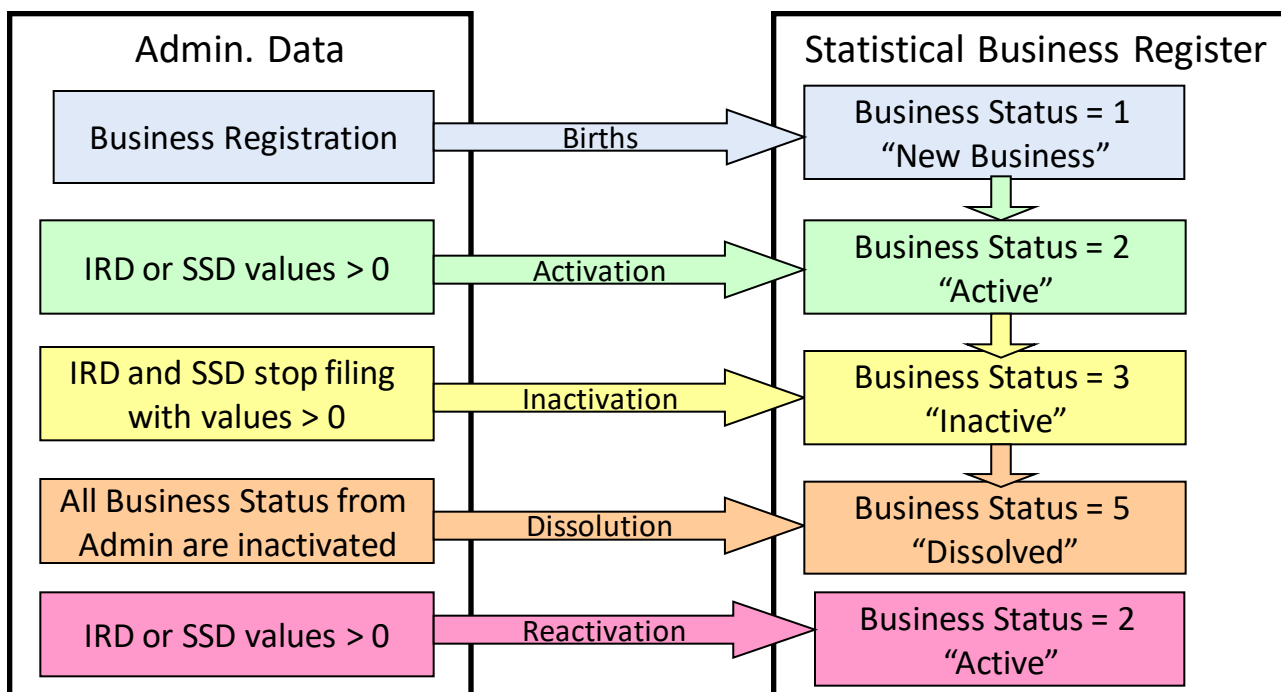
Value added activities from the NSO

- Profiling activities - building Legal/Operating structures
- Harmonization of ISIC and other classifications
- Maintaining a statistical business status
- Creation of Statistical Indicators
- Identification and creation of informal businesses

Administrative Data Integration to the SBR in CARICOM Countries

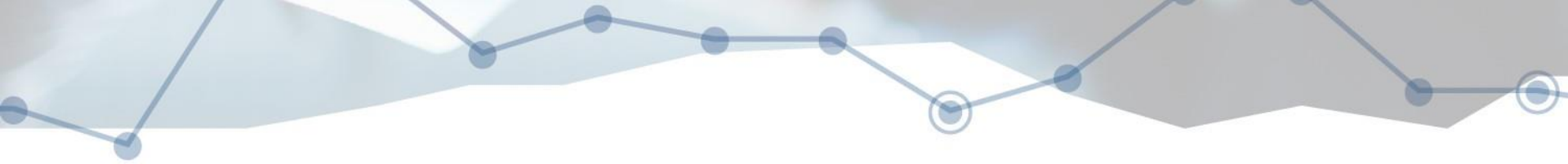


Business Status Flow in CARICOM countries



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





Example



Example (Cont.)

Statistical Business Register (SBR)

| BR Main Table | | | | | | | | |
|---------------|-----------|--------|------|-----|------|-------|---------------|-----------|
| StatID | Bus. Name | Add. | ISIC | Rev | Emp# | Sales | Prospect Flag | Bus. Stat |
| S0001 | ABC | 1 Main | 5510 | | | | 1 | New |
| S0002 | CBC | 2 Main | 5610 | | | | 1 | New |
| S0003 | A&W | 3 Main | 5610 | | | | 1 | New |
| | | | | | | | | |
| | | | | | | | | |

| Link Table | | | | |
|------------|----------|-------|------|-------|
| StatID | REGISTID | TAXID | SSID | VATID |
| S0001 | C1234 | | | |
| S0002 | C2356 | | | |
| S0003 | C4567 | | | |
| | | | | |
| | | | | |



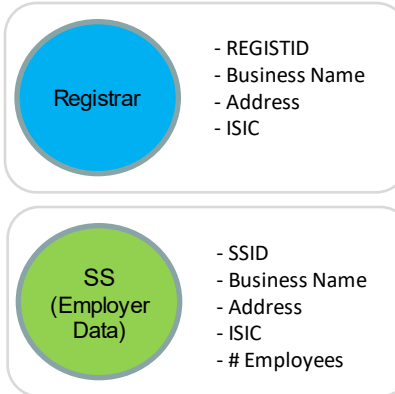
- REGISTID
- Business Name
- Address
- ISIC

Example (Cont.)

Statistical Business Register (SBR)

| BR Main Table | | | | | | | | |
|---------------|-----------|--------|------|-----|------|-------|---------------|-----------|
| StatID | Bus. Name | Add. | ISIC | Rev | Emp# | Sales | Prospect Flag | Bus. Stat |
| S0001 | ABC | 1 Main | 5510 | | 10 | | 0 | Alive |
| S0002 | CBC | 2 Main | 5610 | | | | 1 | New |
| S0003 | A&W | 3 Main | 5610 | | 20 | | 0 | Alive |
| S0004 | XYZ | 5 Main | 1030 | | 65 | | 0 | Alive |

| Link Table | | | | |
|------------|----------|-------|-------|-------|
| StatID | REGISTID | TAXID | SSID | VATID |
| S0001 | C1234 | | N4210 | |
| S0002 | C2356 | | | |
| S0003 | C4567 | | N9365 | |
| S0004 | | | N1367 | |



Example (Cont.)

Statistical Business Register (SBR)

| BR Main Table | | | | | | | | |
|---------------|-----------|--------|------|-------|------|-------|---------------|-----------|
| StatID | Bus. Name | Add. | ISIC | Rev | Emp# | Sales | Prospect Flag | Bus. Stat |
| S0001 | ABC | 1 Main | 5510 | 10000 | 10 | | 0 | Alive |
| S0002 | CBC | 2 Main | 5610 | | | | 1 | New |
| S0003 | A&W | 3 Main | 5610 | 90000 | 20 | | 0 | Alive |
| S0004 | XYZ | 5 Main | 1030 | 22600 | 65 | | 0 | Alive |

| Link Table | | | | |
|------------|----------|-------|-------|-------|
| StatID | REGISTID | TAXID | SSID | VATID |
| S0001 | C1234 | R2001 | N4210 | |
| S0002 | C2356 | | | |
| S0003 | C4567 | R2004 | N9365 | |
| S0004 | | R2005 | N1367 | |

Registrar

- REGISTID
- Business Name
- Address
- ISIC

SS (Employer Data)

- SSID
- Business Name
- Address
- ISIC
- # Employees

IRD

- TAXID
- Business Name
- Address
- ISIC
- Revenue

Example (ont.)

Statistical Business Register (SBR)

| BR Main Table | | | | | | | | |
|---------------|-----------|--------|------|-------|------|--------|---------------|-----------|
| StatID | Bus. Name | Add. | ISIC | Rev | Emp# | Sales | Prospect Flag | Bus. Stat |
| S0001 | ABC | 1 Main | 5510 | 10000 | 10 | 14000 | 0 | Alive |
| S0002 | CBC | 2 Main | 5610 | | | 55555 | 0 | Alive |
| S0003 | A&W | 3 Main | 5610 | 90000 | 20 | | 0 | Alive |
| S0004 | XYZ | 5 Main | 1030 | 22600 | 65 | 38888 | 0 | Alive |
| S0005 | ZZZ | 6 Main | 3555 | | | 150000 | 0 | Alive |

| Link Table | | | | |
|------------|----------|-------|-------|-------|
| StatID | REGISTID | TAXID | SSID | VATID |
| S0001 | C1234 | R2001 | N4210 | V3555 |
| S0002 | C2356 | | | V3666 |
| S0003 | C4567 | R2004 | N9365 | |
| S0004 | | R2005 | N1367 | V3888 |
| S0005 | | | | V3999 |

Registrar

- REGISTID
- Business Name
- Address
- ISIC

SS (Employer Data)

- SSID
- Business Name
- Address
- ISIC
- # Employees

IRD

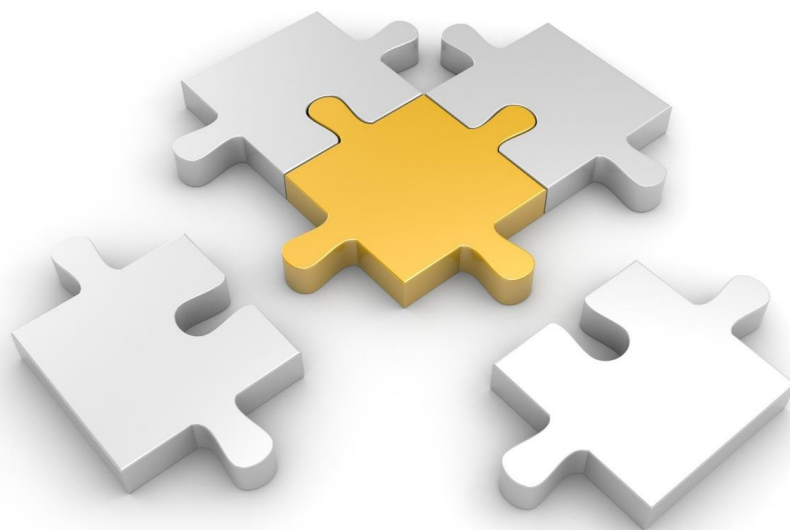
- TAXID (BIRD)
- Business Name
- Address
- ISIC
- Revenue

VAT

- VATID
- Business Name
- Address
- ISIC
- Sales



Data Security





Confidentiality and Security of data

- Administrative data and record linkage offers great opportunity for statistical organisations to create new statistical data products such as statistical registers and other products
- The more data integration that is done, the more sensitive the micro linked database becomes
- Need to take additional security measures to protect linked micro data

Think of creating a process to review, approve and register all record linkage requests



Confidentiality and Security of data (Cont'd)

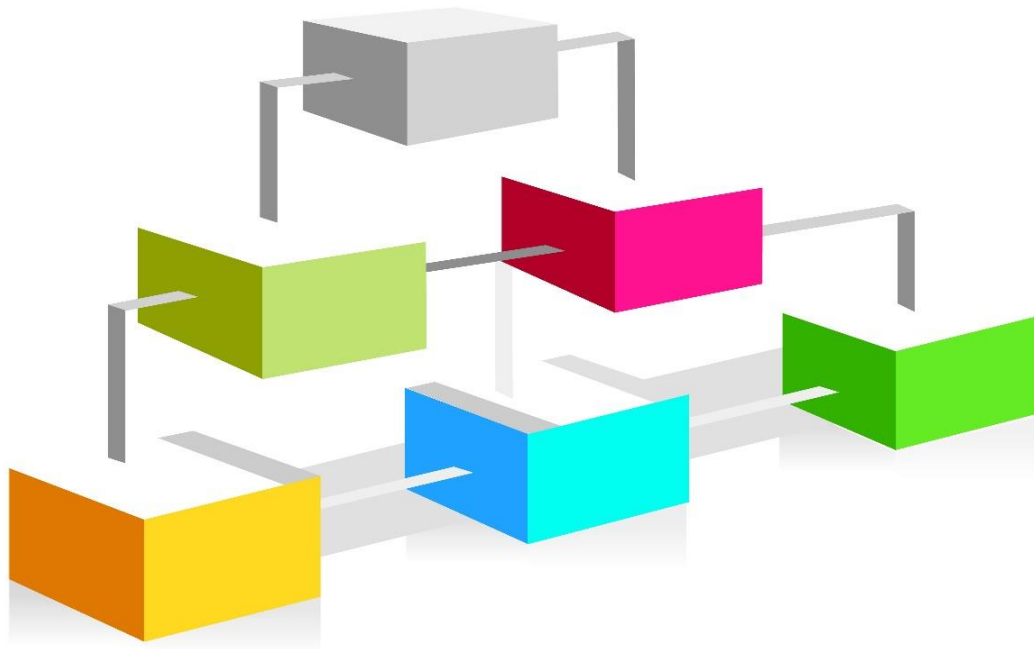
- It is recommended to put security measures in place to protect linked data
 - Control access to the statistical register (SBR) and other linked files
 - Review the list of SBR users on a regular basis (quarterly or yearly) to remove users who no longer require access
 - Store extracted data (SUF) in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



Confidentiality and Security of data (Cont'd)

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

Observations and Recommendations





General Observations

- Often no common unique identifier between administrative data sources (REGISTRAR, SSD, IRD)
- Addresses are not always standardized and post offices are often used
- Telephone numbers are also not always standardized
- Business names are not always reported consistently and sometimes contain:
 - Typography errors
 - Differences between name fields
 - Different record length (truncated fields)



Recommendations

- use a unique business identifier
 - If not possible, try to reuse existing registration numbers such as Registrar-Id, Tax-Payer-Id or Social-Security-ID
- Standardize the addresses
 - Civic number (PO Box), Street Name, Street Type and City in separate fields
- Collect business name identifiers in separate fields
 - Example: Registered Name (Legal) = 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



Advantages of Sharing Data with NSO

- Feedback on data quality
- Access to better official statistics for managing programs and making informed decisions
- Access to a larger variety of cross tabulation aggregate numbers
- Having access to improved industry classification
- Reduce the overall business response burden
- Increase usage of collected data resulting in more efficient usage of public funds
- Facilitate the production of estimates during disaster management
- Harmonization of estimates across the NSS
- Alignment with international practices

38





Conclusion

- Use of micro administrative data by the NSO will help produce reliable official statistics while reducing response burden and costs
- Collaboration between NSO and data providers will result in improving the quality of the administrative data (e.g. Industry classification)
- Facilitate rapid production of statistics during special crisis such as COVID19 and hurricanes
- Continued work and communication with the NSS to identify potential new collaboration initiatives



You can contact the PRASC team at:

prasc@statcan.gc.ca

Canada 