

# PRASC



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

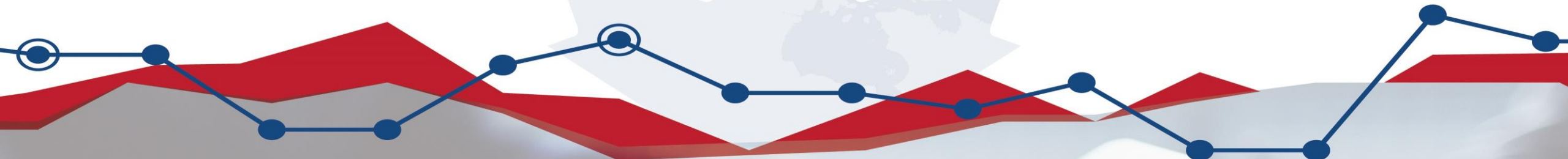
**Projet régional pour  
l'avancement de la statistique  
dans les Caraïbes**

Funded by the  
Government  
of Canada

**Canada**

# Business Rules and Validations

As Presented to the Central Statistical Office of Trinidad & Tobago



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

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



Delivering insight through data, for a better Canada



Statistics  
Canada

Statistique  
Canada

Canada

## Presentation Outline

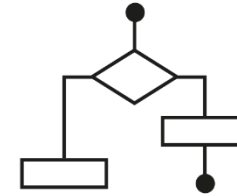
---

- Business Rules and Validations
  - Overview
  - BR Variables
- BR Initialization
- Derivation of BR Variables
- Ongoing Maintenance
- Conclusions

# Business Rules and Validations: Overview

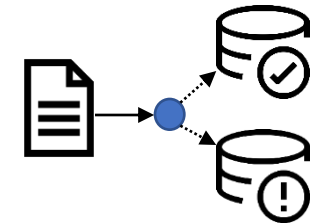
## What are Business Rules?

---



Business rules are statements that describe a criteria or the conditions which determine what can or cannot be done or that result in some action or decision being taken.

E.g. A business can be considered active only if generating size values (revenue/employment).



## What are Validations?

---

Specific rules that can be used to determine the validity of a data value or attribute. Typically applied to input data that can be from any source such as administrative data files, user input, or system created or derived input.

E.g. Data field for ISIC code must be exactly 5 characters in length.

It is common to have an associated business rule that defines the actions or decisions that follow in the event of such a validation failure.

E.g. ISIC code is to be ignored (potentially logged to processing report) and not loaded onto the frame.

## Why have business rules and validations?

---

- Ensure standardization of data and adherence to defined code sets
- Enforce database relationships and coherence between variables
- Set rules and constraints
- Enforce data quality in order to minimize possible issues and problems for sampling, collection and other processes
- Embed, throughout the process the adherence to established concepts
- Avoid unnecessary costs and time!

## Practical Use Cases (1)



Goal: Ensure completeness and correctness of source files before creating / updating BR Frame.



Validation: Compare each source file record layout / variable data types with previous source files received and metadata.

Example: tax identification number must be a numeric format.

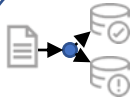


Business rule: if tax identification number fails validation, ignore and log record to processing report and continue to process remaining records.

## Practical Use Cases (2)



Goal: Detection of potential erroneous values or outliers.



Validation: Establish reasonable threshold ranges for certain variables to allow the detection of outliers.

Example: number of employees received from NIB administrative file should be between 1 and 10,000.



Business rule: if data value fails range validations, do not update frame and flag for review.

## Practical Use Cases (3)



Goal: Identify data coherency issues and correct or flag for review.



Validation: Define deterministic and coherence edits that compare values between related fields that should be coherent.

Example: Total number of employees must be equal to or greater than the number of wage earners + number salaried employees.



Business rule: if data value fails deterministic and coherence validations, correct value and update frame (or flag for review).

# Business Rules and Validations: BR Variables

---

## Business Rules and Validations Needs

Determine the types of business rules and validations that are required for the various BR variables and input sources:

- System/Data Integrity Rules or
- Data Quality Assurance Rules or
- Data Coherence Rules or
- No rules required

## System Integrity Rules (hard rules)

- Targeting 100% correctness
- Normally embedded in the system or processes itself
- Examples:
  - VAT Sales amount MUST be a numeric value
  - Size variables MUST have a reference period
  - Business entity MUST only by of a single type of business (Corporation, Sole Proprietor, etc.)

## Quality Assurance Rules (soft rules)

- Targeting a “minimum acceptable” quality level
- Normally not embedded in the system or processes (procedural)
- Examples:
  - VAT Gross Sales amount SHOULD be greater or equal to zero (0) if the business is active
  - A business SHOULD have a complete address
  - A business entity of type ‘corporation’ SHOULD have a corresponding incorporation certificate number
  - An operational unit SHOULD have an address that corresponds to the physical location of where its activities take place (not post office)

## Data Coherence Rules

---

- Targeting data items that should be coherent
- Identify and correct values
- Examples:
  - Number of wage earners or number of salaried employees cannot be greater than the total number of employees
  - Business status and the presence of size values or lack thereof has to be coherent
  - Postal Code and city should agree

# BR Initialization

---

## Initializing the SBR with Admin Data

- Evaluate the sources (Comp. Registry, BID, NIB, others) in terms of their coverage, content and quality
- Determine the most complete, timely and reliable data source for the initial creation of records – Primary Update Source
- Determine other sources that can be used as secondary update sources or linked to Primary source for initialization

## Initialization Steps

---

- Develop rules to determine variable priority by administrative data source
- Similar data items may contain different values
  - Example: Legal Name, Trade Name, Address, ISIC
- Sources can be prioritized depending on:
  - Source reliability
  - Type of update
  - Reference period

## Initialization Steps (cont.)

---

- Perform record linkage to allow business unit to be linked to various Admin sources at birth
- Build Concordance Table
  - Insert a new record into the External Key Table
  - Assign a unique statistical number
- Insert available fields into respective tables

# Derivation of BR Data Variables

---

## Business Status Derivation

---

- Activation
- Inactivation
- Re-activation

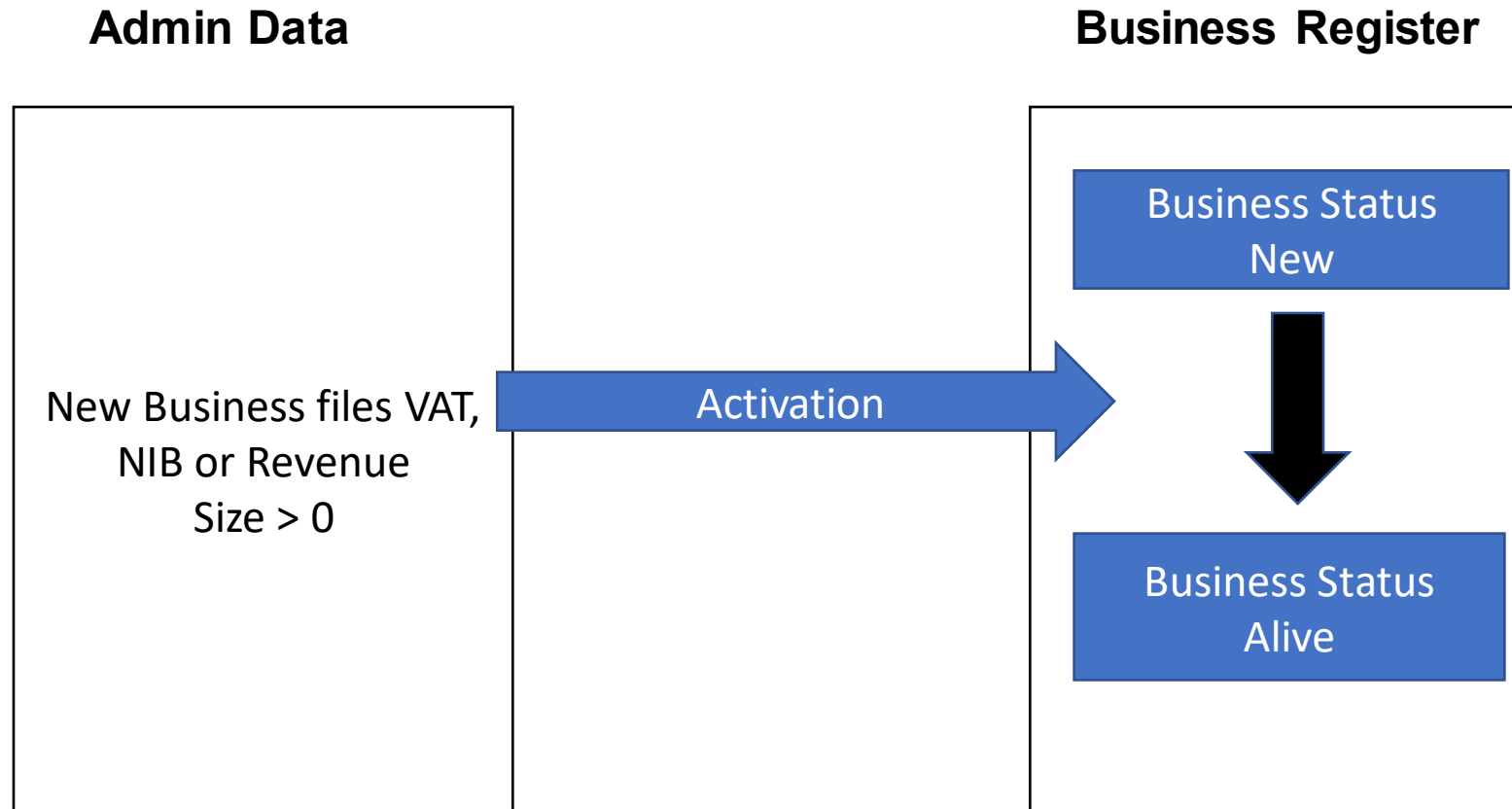
Note: Business status change date is important for the derivation of the Business status in order to apply the Business status change rules

## Business Status Derivation: Activation Rule

- All existing units that have a Business status of “New” and for which there is at least one size variable  $> 0$ 
  - Sales, Employee Info, Revenue, etc.
- OR
- All newly loaded units with at least one size variable  $> 0$

Business Status = Alive

## Activation



## Business Status Rules

---

- Inactivation Rules

- All “Alive” and “New” units with no *recent* remittances

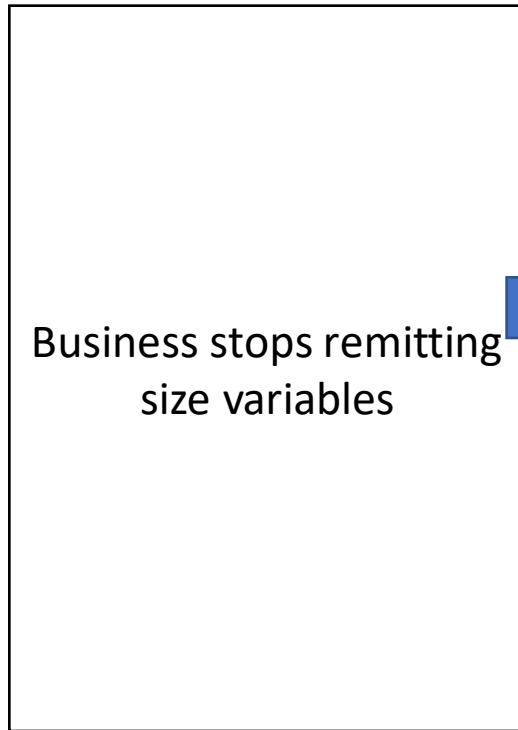
“Recent” defined as\*:

- No revenue (BID) within the last 35 months
- No employment (NIB) within the last 24 months
- No sales (VAT) within the last 13 months

\*NSO can modify time frames to their requirements

# Inactivation

## Admin Data



## Business Register

Business Status  
Alive



Business Status  
Inactive



## Business Status Rules

---

- Reactivation Rules

- Business Status is not “Alive” or “New” and business starts remitting Revenue, Sales or Employment
- Reference period of remittance is more recent than the current business status effective date

## Reactivation

### Admin Data

Newly received files  
VAT, NIB or Revenue  
Size > 0

Re-activation

### Business Register

Business Status  
Inactive/Dissolved



Business Status  
Alive

## **Derived Variables -Recommendations**

- Business status should be derived on a regular basis
  - Prior to the creation of the BR snapshot
- Benefits of more precise coverage
  - Improved quality of estimates
  - Improved response rates
  - Reduced collection costs

## Derived Variables -Recommendations

- Statistical Indicators
  - If Business is not “Alive” OR if ISIC is unassigned, then set Statistical Flags (Enterprise and Establishment) to False
  - If Business “Alive” AND ISIC is assigned, then set Statistical Flags to True (Enterprise or Establishment)\*
    - \*Dependent on the structure complexity

## **Derived Variables -Recommendations**

- Country of Control
  - If populated in Admin Files then set accordingly
  - If not populated in Admin Files, derive from top parent Legal Entity

## Ongoing Maintenance

---

## Variable Priority Table

---

- Used to determine
  - Which variables will be updated from each source
    - Blank – do not include
    - Number – assigns priority importance
  - Sequential order of variable updates by source
- Data Sources include
  - Business Registration
  - Revenue Authority – VAT and Revenue
  - National Insurance
  - Survey
  - Online

# Variable Priority Table Example

VariableName	Data Source					
	BIRPriority	NIBPriority	VATPriority	RegPriority	SelfEmployedPriority	SurveyCollected
Social_Security_Id		1				
REGISTRAR_ID				1		
TAXPAYER_ID	1					
VAT_ID			1			
TEPA_ID						
RegistrationDate	5	4	3	1	2	
FiscalPeriodStart	2	3	4	1	5	
FiscalPeriodEnd	2	3	4	1	5	
BusinessType	2	3	4	1	5	
BusinessStatus	2	3	4	1		
SNAInstitutionalSector	2	3	4			1
ReferencePeriod	2	3	4			1
LegalName	2	3	4			1
LegalStreetAddress	2	3	4			1
LegalTelephoneNumber	2	3	4			1
OperatingName	2	3	4			1
OperatingStreetAddress	2	3	4			1
OperatingTelephoneNumber	2	3	4			1
Website	5	3	4	2		1
ISIC				2		1
Sales			1			2
Revenue			1			2
NumberOfEmployees			1			2
Wages			1			2
NumberOfEmployeesMale						
NumberOfEmployeesFemale						
CountryOfControl						

## Conclusions

---

- Understand admin source data in order to develop Business Rules to protect the integrity of the data on BR
- Determine which updates require hard, soft, deterministic or no business rules
- Initialize BR with the most comprehensive and reliable information source – Primary Data Source + Secondary sources
- Develop a maintenance schedule and procedures for on-going updates to BR
- Continue to improve the record linkage and business rules processes

You can contact the PRASC team at:

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

Canada