SHARING ON MALAYSIA EXPERIENCE:
INTEGRATION OF MBSR WITH TRADE DATABASE

RIYANTI BINTI SAARI
METHODOLOGY & RESEARCH DIVISION

29-31 OKTOBER 2018
Technical Mission of SESRIC An Business And Registers,
Brunei Darussalam
OUTLINES

- STATSBDATION JOURNEY
- STATSBDATION ARCHITECTURE
- STATSBDATION MODULES
- WHAT IS TEC?
- INTERNATIONAL RECOMMENDATION
- TEC OUTPUTS
- TEC OUTCOMES
- PORTAL
Inception Meeting in Colombo, Sri Lanka
30 June - 2 July 2014
Organizer: ADB

CDTA: 8594 - Statistical Business Registers for Improved Information on Small Medium-Sized and Large Enterprises.

MSBR rich with businesses information & become richer when it is combined with trade database.
STATSBDA MODULES

Trade by Enterprise Characteristic (TEC)

Price Intelligence (PI)

Real Time Business Status (RTBS)

BizCode@Stats (Mobile App)

Public Maturity Assessment on Official Statistics (PMAOS)

Real Time News on Official Statistics (RTOS)
STATSBDA MODULES

- Trade by Enterprise Characteristics
- Price Intelligence
- Public Maturity Assessment of Official Statistics
- Real Time Business Status
- Bizcode@Stats
- Real Time News of Official Statistics
High volume of administrative data from JKDM is integrated with MSBR to provide value added to an existing international trade statistics without having to run a new survey.

These statistics are able to identify the characteristics of companies or businesses involved global market.
INTEGRATION WITH EXTERNAL TRADE REGISTER

Para 2.54:
Coherent compilation of trade statistics by enterprise characteristics requires linkage of trade and business registers at micro level. If this can be achieved, the combination of the key enterprise characteristics and the trade data, such as product code and partner country, offers many opportunities for producing a more complete and diversified view of the structure of both trade and production.

LINKING TRADE AND BUSINESS STATISTICS

Para 11.5:
Linking and integrating trade and business statistics is important for data-compilation and analytical purposes.

A major development in economics statistics in recent years has been the establishment and use of national business registers that allow not only the conduct of business surveys but also the linkage of information from different data sources, potentially leading to significant efficiency and quality gains in data collection.
The UN Statistical Commission recommends national statistical offices to adopt an integrated approach to economic statistics in order to attain:

- **Consistency** in measuring economic activity across sectors and regions
- **Global consistency** of key economic indicators
- **Greater accuracy** by reconciling discrepancies among data sources
- **Reduction collection and reporting burden** by using the same information for different purposes
One effective and proven application is linking the SBR to trade statistics to compile **trade by enterprise characteristics (TEC)**, which can be compiled for merchandise trade, trade in services, & foreign direct investment.

TEC data aim to describe trade flows from the viewpoint of enterprises:

- **Impact of international trade on employment, growth and income** (by kind of industry, size of enterprise and other characteristics)
- **Effectiveness of trade policies** (e.g., export-promotion, effects on 2-way traders and foreign affiliates)
Trade Database
- ~1m records
- ~11 files with different column names
- ~80 columns

MSBR
- ~3m records
- ~2 files
- ~200 columns
Rich with businesses information & become richer when it is combined with other databases

MSBR + Trade Database

Big Data platform is used to process the data within a stipulated time

Big Data Revolution

New information that is not exist in stand-alone statistical domains

Data Management

Quick data quality checks and analysis

The effort is less efficient with existing business solutions

MSBR

Rich with businesses information & become richer when it is combined with other databases

TEC Statistics

New information that is not exist in stand-alone statistical domains
TRADE DATABASE

- Business Registration Number
- Business Name
- Business Address
- Product Exported/Imported
- Partner Country
- Exports/Imports Value
- Exports/Imports Volume
- etc.

MALAYSIA STATISTICAL BUSINESS REGISTER

- Business Registration Number
- Business Name
- Business Address
- Industry
- Incorporated Date
- Operating Status
- Number of Employees
- SME Status
- etc.

TRADE BY ENTERPRISE CHARACTERISTICS DATABASE

- Business Reg. No.
- Business Name
- Business Address
- Product Exported/Imported
- Partner Country
- Exports/Imports Value
- Exports/Imports Volume
- Industry
- Incorporated Date
- Operating Status
- Number of Employees
- SME Status
- etc.
In an ideal scenario.....

 Trade Database  MSBR

Business Registration Number (BRN)

MATCH
Due to poor quality of BRN information....

**Text Matching**
- Term Extraction
- Term Weighting
- Algorithm & Measure
- Scoring

Score: 88% considered as matched
TEC PROCESS FLOW

Raw Data Profiling & Import

Lookup Data Cleansing & Standardization

MSBR Cleansing, Standardization & Matching

Trade Cleansing, Standardization & Matching

MSBR & Trade Aggregation

Derivation

DB2 TEC Extraction

TEC Integration (MYSQL-DB2)

TEC Indexing & Dictionary quality improvement

Post-processing & Generate TEC

Aggregation

Lookup Table Preparation

Aggregated TEC Lookup

TEC Integration (SAS-DB2)

Visualization (Tableau)
**TEC Process Flow (Details)**

### Raw data profiling & import
- Import raw data from flat files format into database
- Update field name and standardize the format
- Verify control figure

### Lookup data cleansing & standardization
- Update field name and standardize the format
- Process lookup data which involve data cleansing and matching
- Logstash lookup data into elasticsearch
- Profile lookup data matching

### MSBR cleansing, standardization & matching
- Update field name and standardize the format
- Process MSBR data which involve data cleansing and matching
- Logstash MSBR data into elasticsearch
- Profile MSBR data matching
<table>
<thead>
<tr>
<th>Trade cleansing, standardization &amp; matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update field name and standardize the format</td>
</tr>
</tbody>
</table>

**MSBR & Trade Aggregation**

Aggregating data for use cases that require data to be aggregated

**Post-processing & Generate TEC**

Create script to join MSBR and Trade data as TEC by month  
Insert TEC data into one table for one year data  
Update selected field in Database to standardize the format
TEC Process Flow (Details) cont’d

**TEC Indexing**
- Create configuration file to logstash
- TEC data into elastic search
- Profile the data

**TEC Integration**
- Load TEC data into DB2
- Verify control figure
**TEC Process Flow (Details) cont’d**

**DB2 TEC Extraction**
- Extract new TEC monthly data for data processing purpose.

**Aggregation**
- Aggregate TEC data into 3 set of aggregated datasets
  - TEC_MAIN
  - TEC_PROD
  - TEC_COMP

**Derivation**
- Derive needed variables for visualization purpose.

**Lookup Table Preparation**
- Extract necessary lookup variables from each lookup table

**Aggregated TEC Lookup**
- Extract necessary lookup variables from each lookup table

**TEC Integration**
- Load 3 final aggregated datasets into DB2 for visualization.
TEC – OUTPUT

• Descriptive
  ➢ Combination of Trade and Company Info (Trade Enterprise Characteristic)
  ➢ Able to deep dive to more granular level
  ➢ Benefit for organization to enhance productivity: Internal & Explicit stakeholders, Public

• Forecasting
  ➢ Gain ability to forecast
  ➢ Maximize the potential of the products or services produced or offered by Malaysia
  ➢ Have an expectation for next period of trade value
  ➢ Alert users when trend is out of expectation and DOSM enable to find the reasons at early stage
**TEC – OUTPUT**

**Descriptive**
- Overall
- Product Code
- Major Trading Partners
- State
- FTA & Non-FTA Trading Partners Tag
- Two-Ways Trader Statistics
- Border State Statistics
- High-Tech Product Statistics
- Broad Economic Category Statistics
- Halal Products Export
- Top Exporter / Importer by Product
- Product by Importer Exporter
- Business Status
- Trade Statistics by Sector
- Trade Statistics by Section (MSIC)
- Economic Region Statistics
- Revenue & Employment Size Band Statistics
- Manufacturing Statistics
- Enterprise Age Statistics
- High Demand Product Statistics

**Forecasting**
- Forecast Total Import/Export Value by Sector for Overall and Among SME
- Forecast Import/Export Value for Each State
- Forecast Number of SME Importers/Exporters by Sector
- Forecast Total Import/Export Value by Top 10 Trading Partners
- Forecast Retained Import Value for Each Broad Economic End Use Category
- Forecast Import/Export Value for Each State
- Forecast Total Import/Export Value of FTA Partner Country
- Sustainability of Top 5
- Forecast the Percentage Contribution in Export Value of Overall SME and Import & Export Value of Overall MNC Respectively at Monthly Level
- Forecast the Overall Import & Export Trade Value at Monthly Level
Trade Statistic by Sector

Export & Import Value by MSIC Sector

Value by Month & Year of Selected Sector
WELCOME TO STATSBDA PORTAL

Powerful Data. Valuable Result. Informative Decision

The Department of Statistics Malaysia (DOSM) is now moving towards Big Data Analytics (BDA) in producing more reliable and timeliness official statistics. Thus, DOSM has initiated the implementation of BDA with the project of STATSBDA which is in line with DOSM’s Transformation Plan 2015-2020. The objectives of the implementation of STATSBDA are:

statsbda.dosm.gov.my
TRADE BY ENTERPRISE CHARACTERISTIC (TEC)

Setting the foundation for the harmonization of economic statistics, the Malaysia Statistical Business Register (MSBR) has been designed, developed and enhanced by Department of Statistics Malaysia (DOSM). MSBR is a fundamental property in maintaining the comprehensive list of businesses and companies operating in Malaysia. It is used as the source of frames for all business surveys. MSBR plays an essential role in the construction and maintenance of an integrated economic information system. The growing demand for better and more detailed business statistics has put focus on MSBR and its role in the production of business statistics.

The function of MSBR is further expanded by integrating it with trade database to be able the trade information connects to the traders’ characteristics. By integrating these two databases, DOSM can be in the better position to provide statistical indicators on trade by enterprise characteristics. Micro-data linking of MSBR and trade is to gain more data insights without initiating new survey. Ultimately, it enriches the international trade statistics by providing closer views of traders who are actively engaged in importing and exporting. This project is conducted to complement conventional trade statistics for an in-depth trade statistics of which the traditional trade statistics are unable to portray its business characteristic, economic activity, revenue size, enterprise age etc.
Strategic partnership with SSM through online system integration between DOSM and SSM which is known as SSM Middleware. It is a web service where DOSM needs to send requests based on specified format to get latest businesses’ / companies’ information.
Checking business status centrally

Allow DOSM to get latest business status registered with SSM through SSM Middleware

SSM Middleware provides a platform for accessing information DOSM business entities registered with SSM

The information received has the characteristics of integrity that is timely, accurate and reliable.

Reduce manual work processes where business information requests are made in batch form