









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



- STATSBDA JOURNEY
- STATSBDA ARCHITECTURE
- STATSBDA MODULES
- **WHAT IS TEC?**
- **INTERNATIONAL RECOMMENDATION**
- **TEC OUTPUTS**
- TEC OUTCOMES
- PORTAL

STATSBDA JOURNEY





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.

TRADE BY ENTERPRISE **CHARACTERISTICS (TEC)**



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



Seminar on Integrated Trade Database with MSBR: A Big Data Initiative

17 September 2015 Sponsored by: ADB Consultant: Mr. Markie Muryawan (Chief International Merchandise, Trade Statistics Section, UNSD)







STATSBDA Kickoff 2 December 2016







Structured Data (TEC)



Embedded in STATSBDA

STATSBDA

Project

Project - decide to harness the massive amounts of trade data using big data platform.



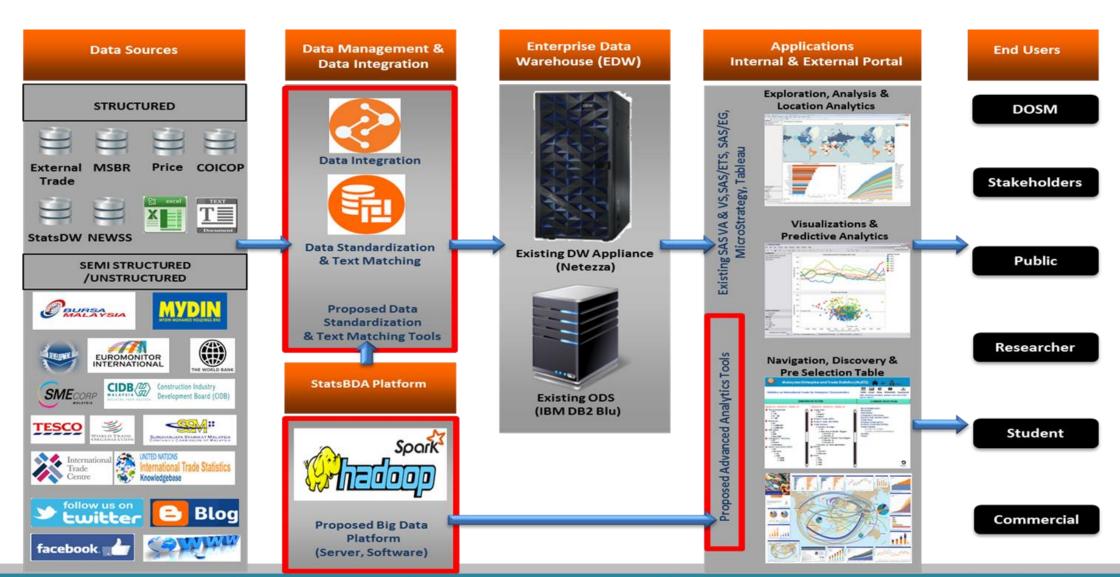






STATSBDA ARCHITECTURE







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)



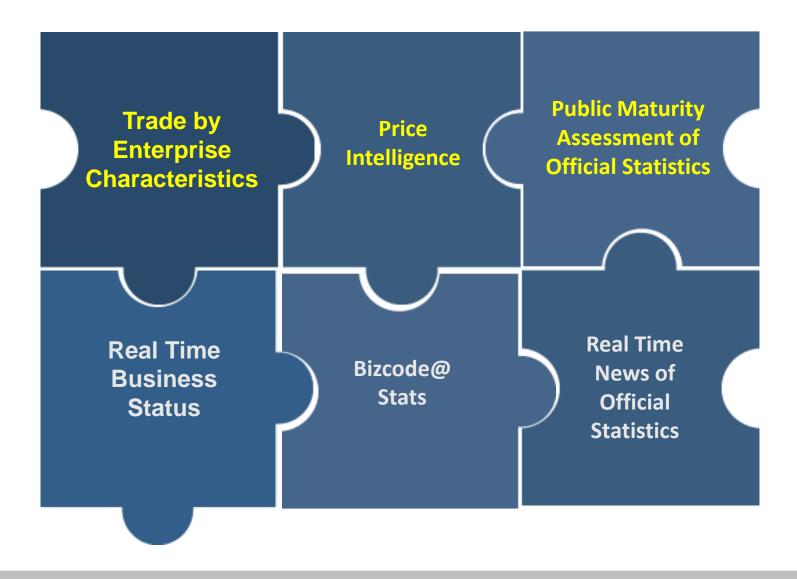






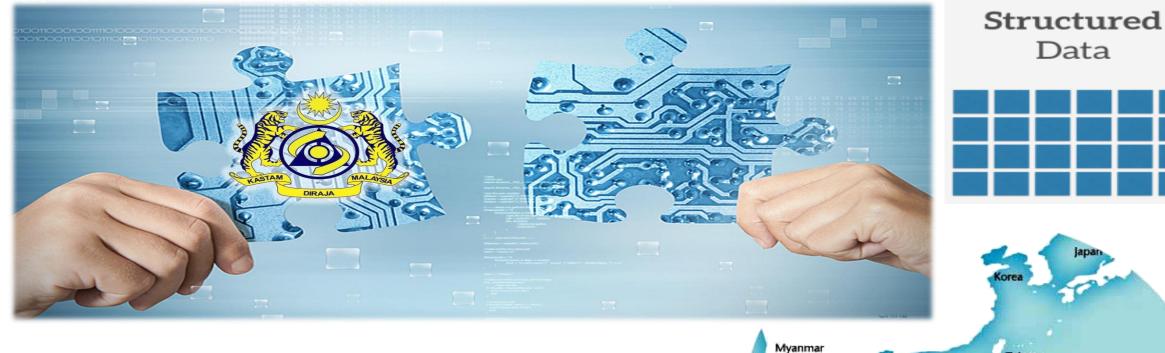
STATSBDA MODULES





TRADE BY ENTERPRISE CHARACTERISTICS (TEC)





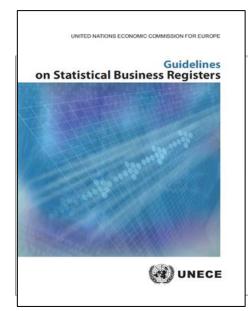
High volume of administrative data from JKDM is integrated with MSBR to provide value added to an existing international trade statistics without having to run a new survey.

These statistics are able to identify the characteristics of companies or businesses involved global market.



INTERNATIONAL RECOMMENDATION





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.



INTERNATIONAL RECOMMENDATION



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

INTERNATIONAL RECOMMENDATION



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 view point 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., exportpromotion, effects on 2-way traders and foreign affiliates)

TEC





Trade Database

- ~1m records
- ~11 files with different column names
- ~80 columns



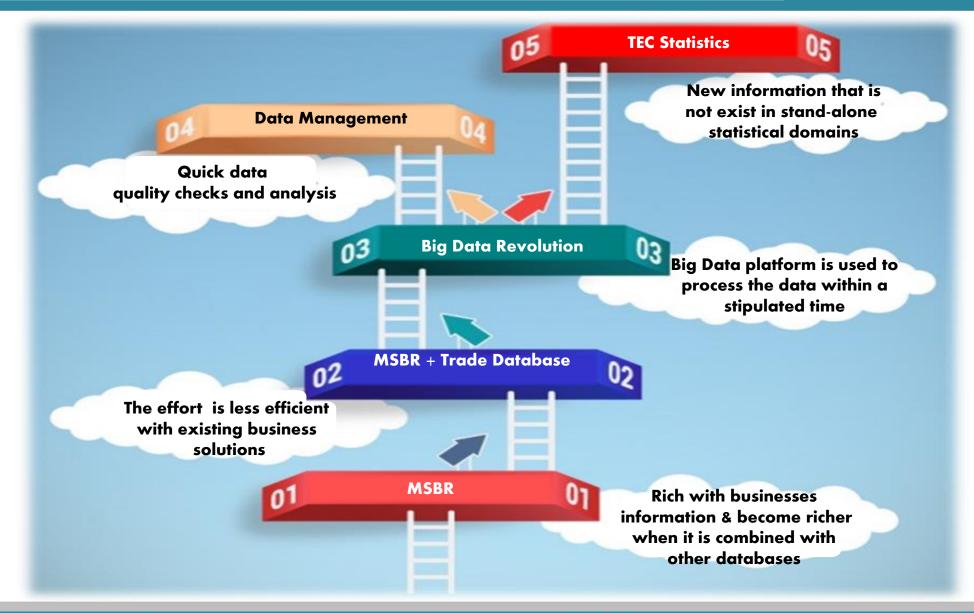
MONTHLY DATA

MSBR

- ~3m records
- ~2 files
- ~200 columns

TEC



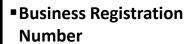


TEC

TRADE DATABASE







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



TRADE BY ENTERPRISE **CHARACTERISTICS DATABASE**



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



- Business Name
- Business Address
- Product Exported/ **Imported**
- Partner Country

- **■** Exports/Imports Volume
- Industry
- Incorporated Date
- Operating Status
- Number of Employees
- **■SME Status**

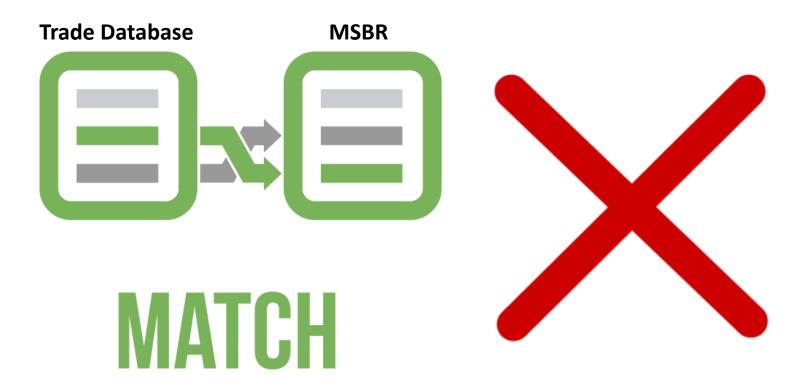




TEC PROCESS FLOW



In an ideal scenario.....

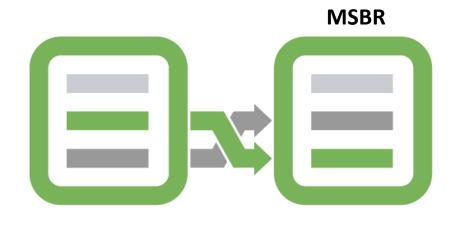


Business Registration Number (BRN)

TEC PROCESS FLOW



Due to poor quality of BRN information....













- Data cleansing
- Data standardization



Text 2

- Data cleansing
- Data standardization



Text Matching

- Term Extraction
- Term Weighting
- Algorithm & Measure
- Scoring

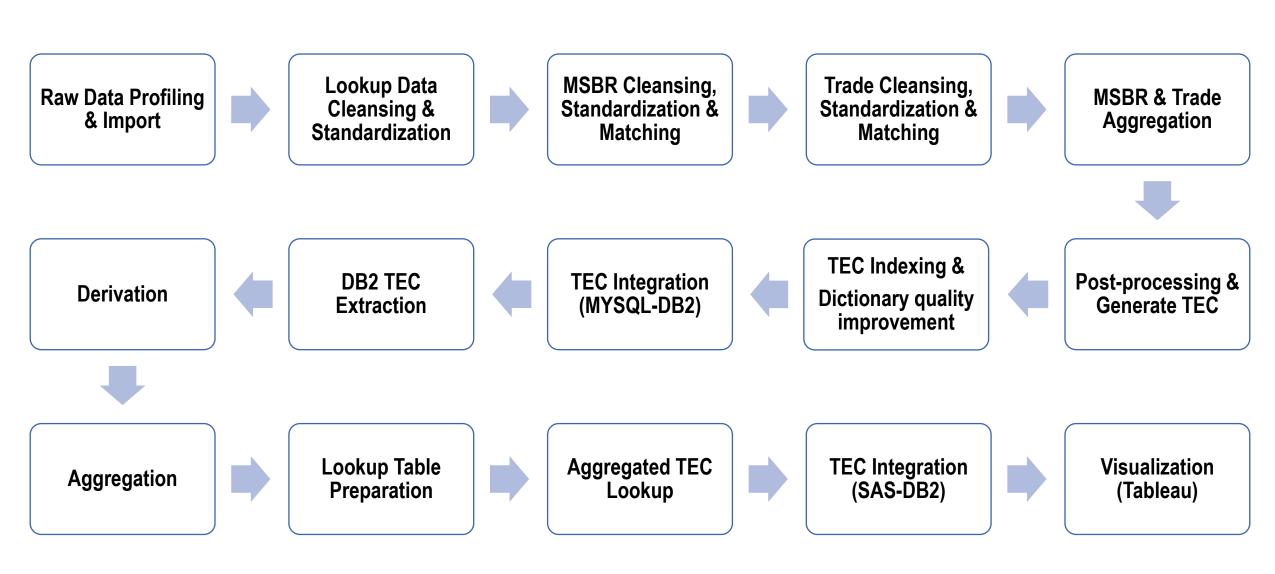


Score:

88% considered as matched

TEC PROCESS FLOW





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





TEC Process Flow (Details) cont'd



Trade cleansing, standardization & matching

Update field name and standardize the format Process trade data which involve data cleansing and matching

Profile trade data matching



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

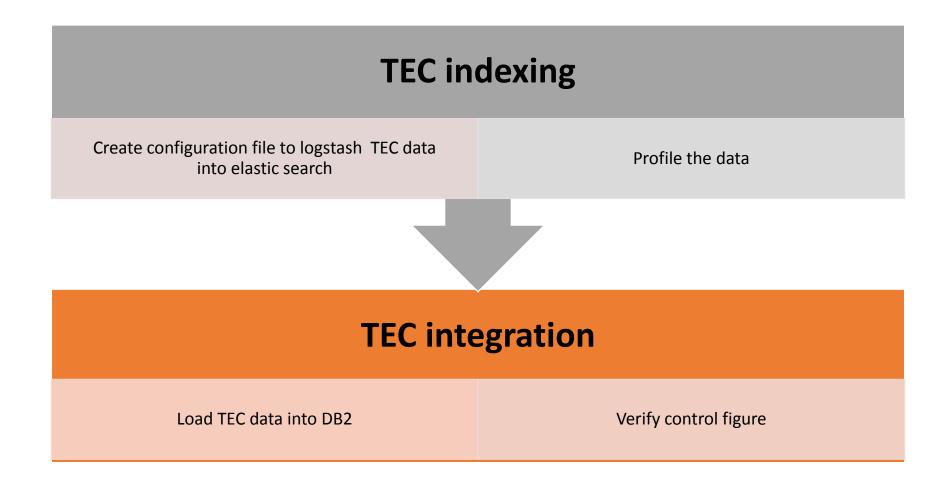
Update selected field in Database to standardize the format





TEC Process Flow (Details) cont'd





TEC Process Flow (Details) cont'd



DB2 TEC Extraction

Extract new TEC monthly data for data processing purpose.



Derivation

Derive needed variables for visualization purpose.



Aggregation

Aggregate TEC data into 3 set of aggregated datasets

TEC_MAIN

TEC_PROD

TEC COMP

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

TEC - PUBLIC USER











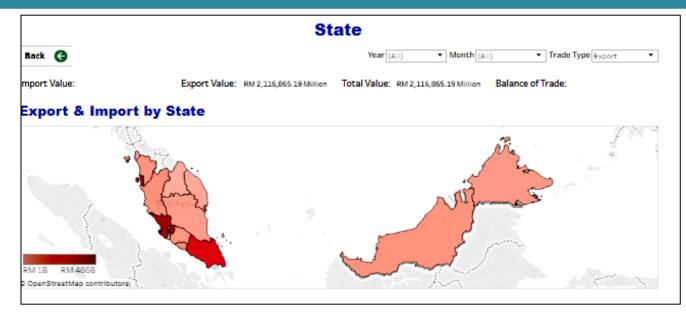


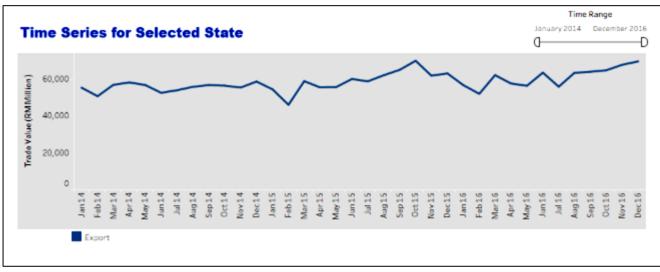


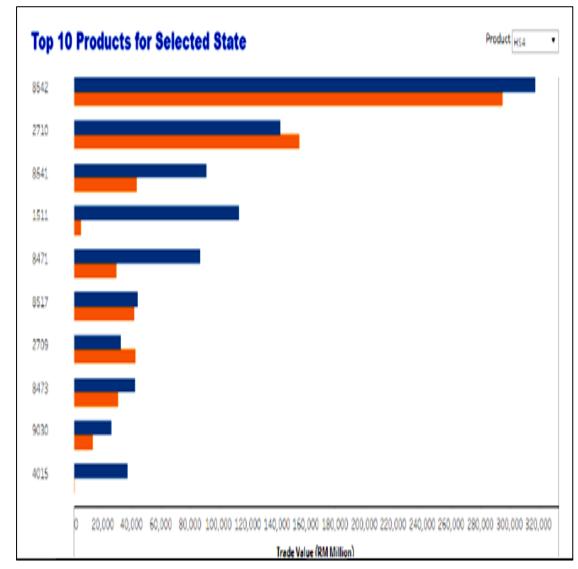


TEC - EXAMPLE









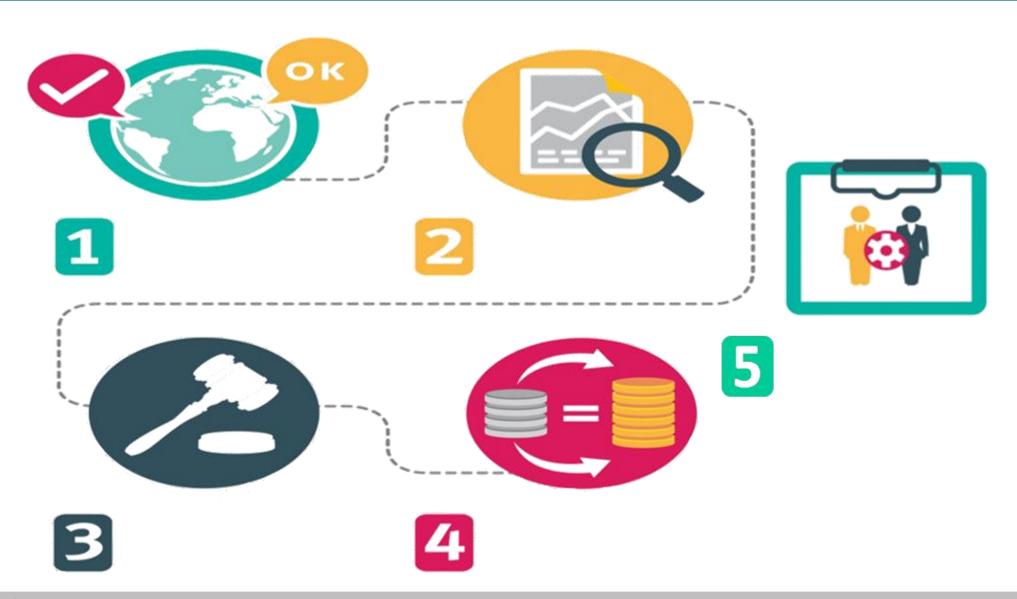
TEC - EXAMPLE





OUTCOME





PORTAL





WELCOME TO STATSBDA PORTAL

Powerful Data, Valuable Result, Informative Decision

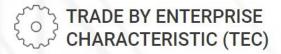
The Department of Statistics Malaysia (DOSM) is now moving towards Big Data Analytics (BDA) in producing more reliable and timeliness official statistics. Thus, DOSM has initiated the implementation of BDA with the project of STATSBDA which is in line with DOSM's Transformation Plan 2015-2020. The objectives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

PORTAL





HOME MODULES NEWSLETTER DATA SCIENCE LAB GALLERY CONTACT US



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.





Explore







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.



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

Checking business status centrally

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

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





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





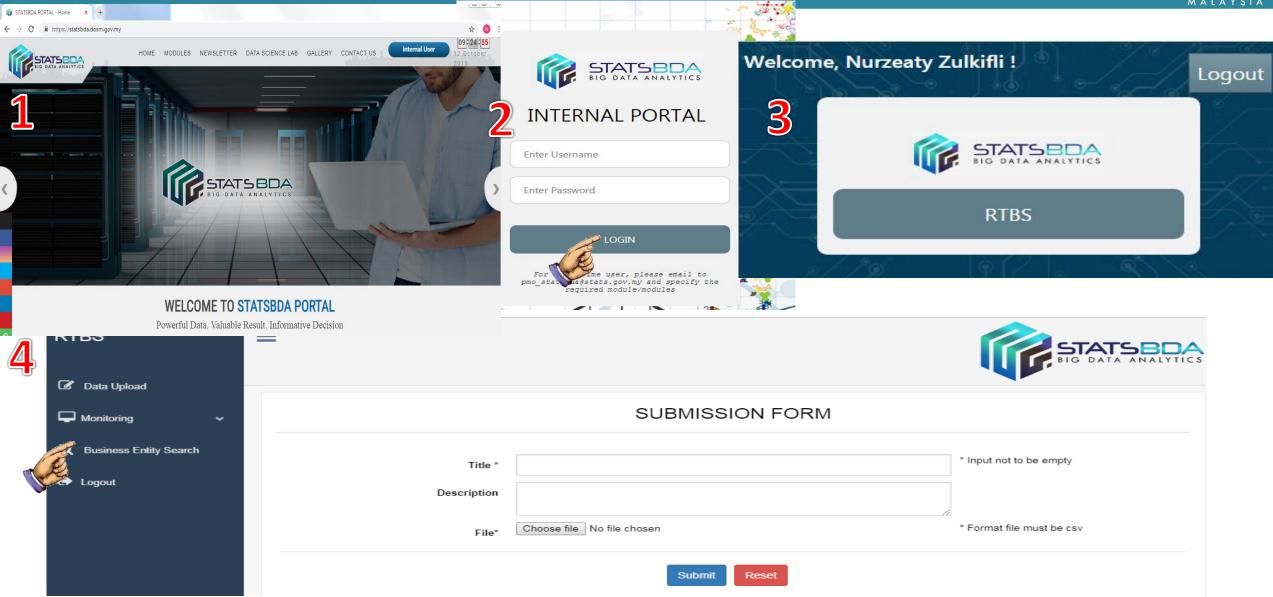


Jabatan Peranakaan

MALAYSIA











TERIMA KASIR & TRANK YOU











#StatsMalaysia | #MyStatsDay | #isiwsc2019 | #MyCensus2020