



# Webinar on Building a Central Statistical Data Warehouse Doha, 17 May 2022

حلقة عمل عن بعد بشأن بناء قاعدة بيانات إحصائية مركزية  
الدوحة، 17 مايو 2022



PRIME MINISTER'S DEPARTMENT  
DEPARTMENT OF STATISTICS MALAYSIA

**StatsDW**<sup>®</sup>

Statistics Data Warehouse

*Brings Data To Life*

# MODERNIZING DATA GOVERNING: StatsDW AS NATIONAL AGENDA

**17<sup>th</sup> MAY 2022 | 2.00 p.m**

**QATAR**

---

Data Management & Integration Division  
Department of Statistics Malaysia



     
StatsMalaysia  
[www.DOSM.gov.my](http://www.DOSM.gov.my)



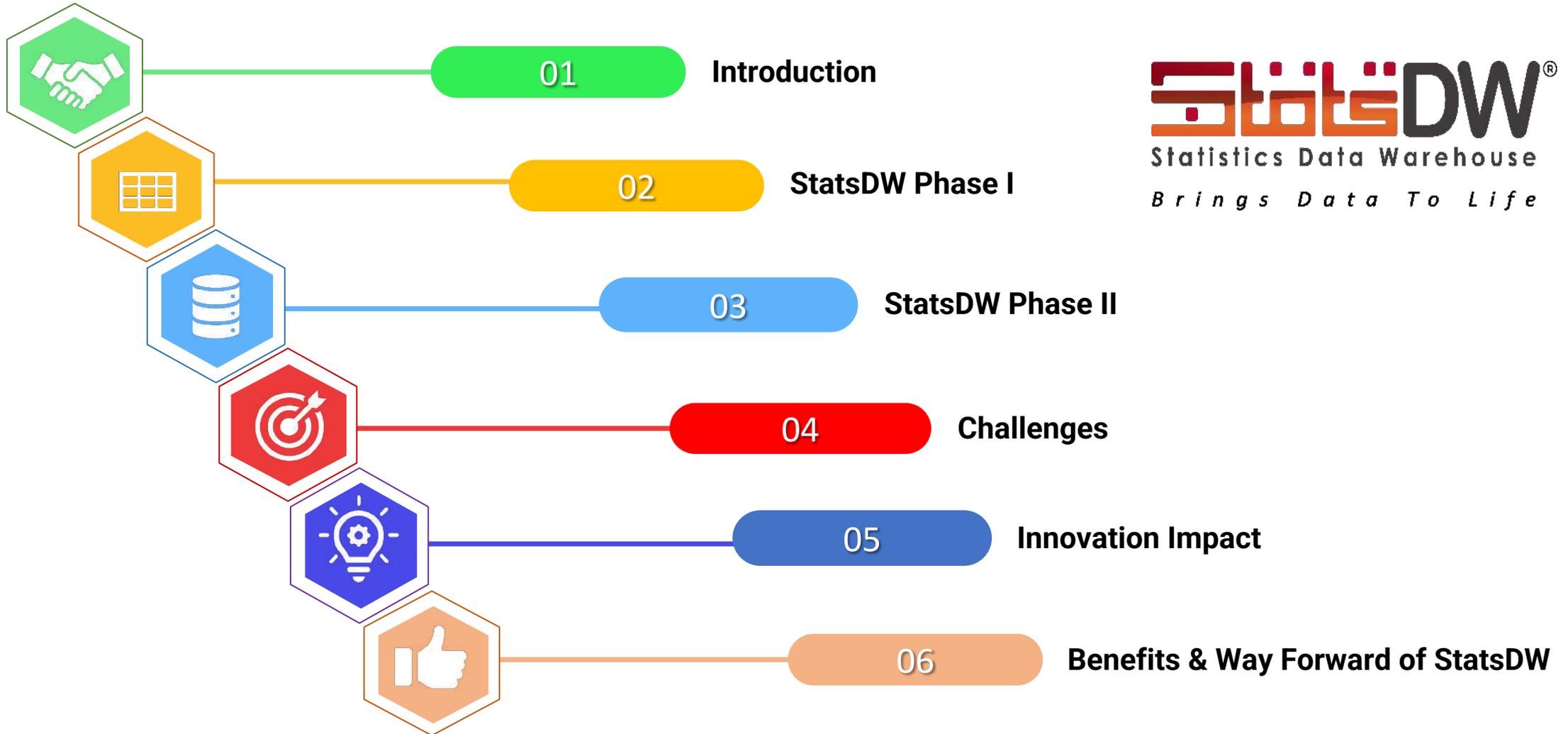
20 OCT



2016 - 2030



# CONTENTS



**StatsDW**<sup>®</sup>  
Statistics Data Warehouse  
*Brings Data To Life*



# INTRODUCTION



     
StatsMalaysia  
[www.DOSM.gov.my](http://www.DOSM.gov.my)



 #KELUARGA  
MALAYSIA



# STRENGTHENING OF NATIONAL STATISTICAL SYSTEM



Efforts to **continuous** modernisation and innovation in order to provide effective service delivery.

## Overflow of Data from Various Sources

1. Data is a highly valued commodity that is widely available across agencies and platforms.
2. Requires efficient and integrated data resource management.
3. A system framework is essential for holistic statistical production.

## Technology Revolution

The need for complex and granular information down to the smallest geographical level and based on various specific demographic characteristics.

## Data requests become more complex

1. Supports the transition to a more efficient and effective high-volume data integration and coordination platform.
2. In line with the digital government and public data era, for better and more effective and cost-effective data processing, structuring, analysis, and handling.

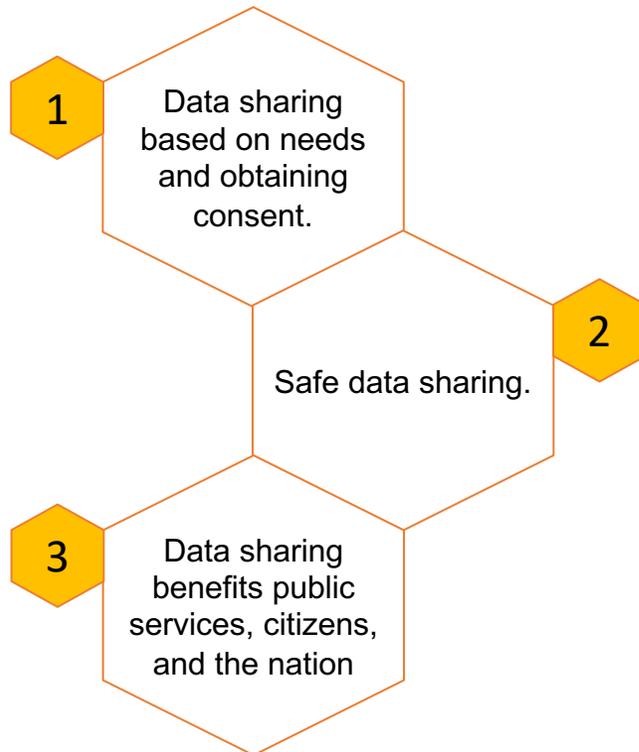
## Environmental change

1. The global situation's uncertainties, and the need for readiness in the preparation of the latest inputs.
2. Rapid economic evolution and business models.
3. Social and environmental management are becoming more complex.



The Public Sector Data Sharing Policy (DPDSA) was clarified in the Public Administration Development Circular number 2 year 2021 Public Sector Data Sharing Policy, which became effective on 27<sup>th</sup> August 2021.

## Principles of Public Sector Data Sharing



## Compliance in Data Sharing





**Statistics Data Warehouse (StatsDW)** is an Enterprise Data Warehouse (EDW) containing **historical data** from various censuses and surveys conducted by the Department of Statistics Malaysia (DOSM) as well as compiled data for economic indicators.



StatsDW **enable fast and easy access to data** and facilitate wider access for users according to user requirements and level of access permitted.



“StatsDW is the 1<sup>st</sup> initiative by the DOSM to consolidate & integrate all the Statistical Data in one **Enterprise Data Warehouse (EDW)**”

**EDW** consolidates data from multiple sources, giving the right people access to the right information so that they can take necessary action. In turn, this helps in reporting, data analysis and acts as a core component of business intelligence/analytics environment.



# OBJECTIVES



01

To consolidate all historical micro and aggregate data in an Enterprise Data Warehouse (EDW)



02

To enhance the quality and accuracy of data

03

To enable fast and easy access to data repository



04

To disseminate data to wider users in timely manner





# MOTIVATION



1

## DATA MANAGEMENT

- Various sources and form of data
- Management and compilation of time series data
- Data Migration



Data is stored in 66 silo /stand alone databases

2

## LIMITED DATA WAREHOUSE APPLIANCES

- Business Intelligence (BI)
- Extract Transfer and Loading (ETL)
- BI Integration vs Geographical Information
- Data Warehouse



Approximately 7,700 softcopies / hardcopies of publication are stored at various location



3

## SYSTEM CAPABILITY

- System integration not available
- Data access and preparation
- Limited data storage



Approximately 5,000 micro data are stored separately at various Subject Matter Divisions

4

## SECURITY AND USER PROFILE

- Non-centered data storage location
- Difficulties in controlling data access





# BENCHMARK STUDY

PRIME MINISTER'S DEPARTMENT  
DEPARTMENT OF STATISTICS MALAYSIA



     
StatsMalaysia  
www.DOSM.gov.my



## STATISTICS CANADA



### CANSIM

- Data warehouse
- Data management
- (<http://www5.statcan.gc.ca>)

## STATISTICS NETHERLANDS



### StatLine

- Micro data management
- Aggregate data management
- eDataBank
- Visualisation
- (<http://statline.cbs.nl>)

## EUROSTATS

- Data warehouse
- Micro data management
- Aggregate data management
- (<http://ec.europa.eu/eurostat/web/europeanstatistical-system>)

## STATISTICS KOREA

### Korean Information Statistical Service (KOSIS)

- Micro data management
- eDataBank
- (<https://kosis.kr/eng/>)



## AUSTRALIAN BUREAU OF STATISTICS

### ABS.Stat

- Data warehouse
- MyLab
- (<http://stat.data.abs.gov.au/>)



## STATISTICS NEW ZEALAND

### NZ.Stat

- Data warehouse
- MyLab
- Visualisation
- (<http://nzdotstat.stats.govt.nz>)

## Guiding Principles

(<https://sqa.mampu.gov.my/index.php/ms/garis-panduan/garis-panduan-pengurusan-projek-ict-prisa>)





# StatsDW JOURNEY

PRIME MINISTER'S DEPARTMENT  
DEPARTMENT OF STATISTICS MALAYSIA



  
StatsMalaysia  
www.DOSM.gov.my



## PRE-DEVELOPMENT

## DEVELOPMENT-PHASE 1

## DEVELOPMENT-PHASE II

2009

2012

2013

2014

2015

2016

2017

2018

2019

2021

Enterprise data warehouse study

Report of study

- Technical needs
- Business needs
- Financials

Development of StatsDW

- Go Live 2016
- 7 Modules

Study for enhancement

- POC
- POT
- Paper work

Development of StatsDW 2.0

- Expansion of subjects
- System Integration within DOSM & OGAs



# StatsDW PHASE I

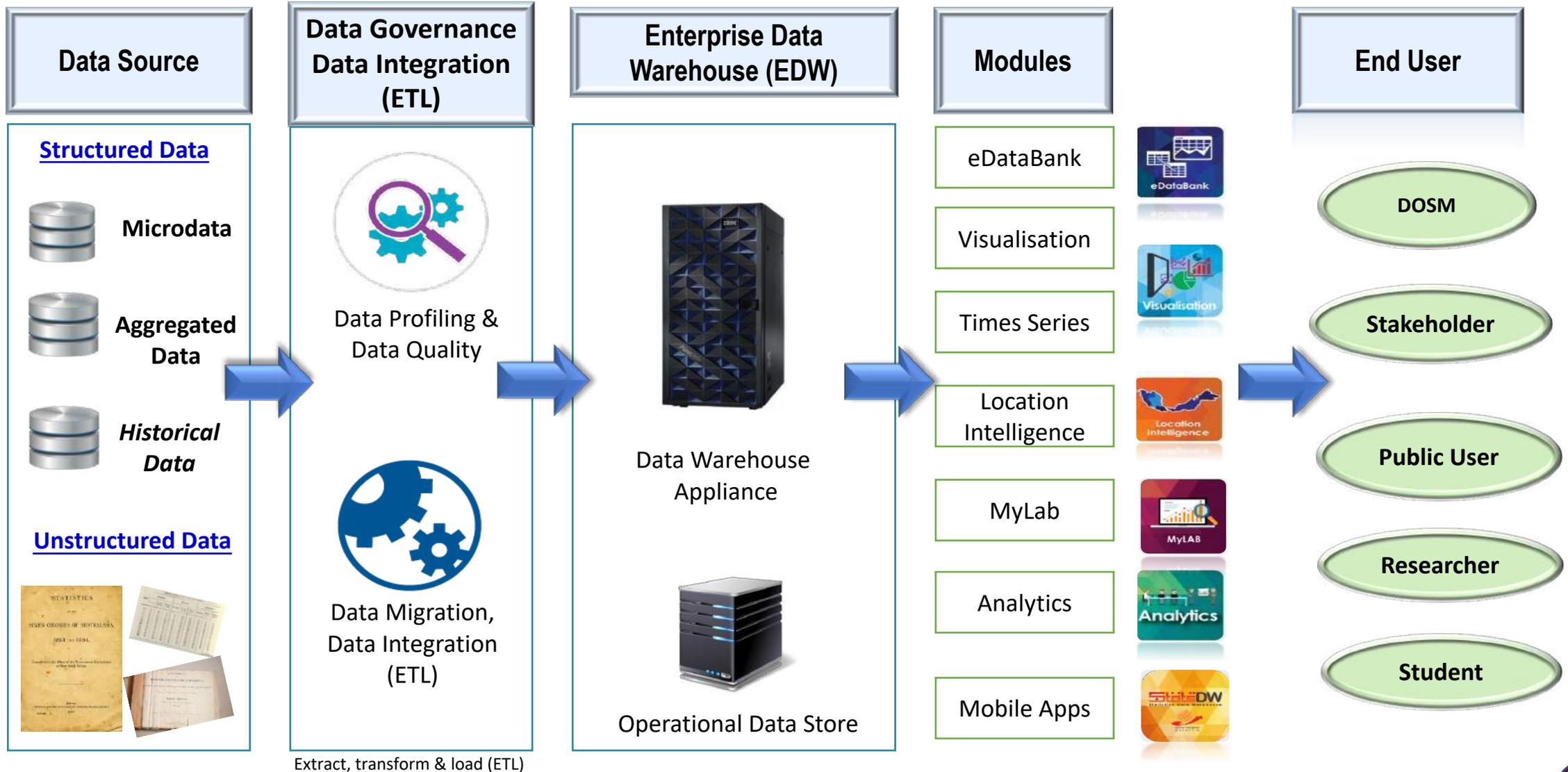


  
StatsMalaysia  
[www.DOSM.gov.my](http://www.DOSM.gov.my)



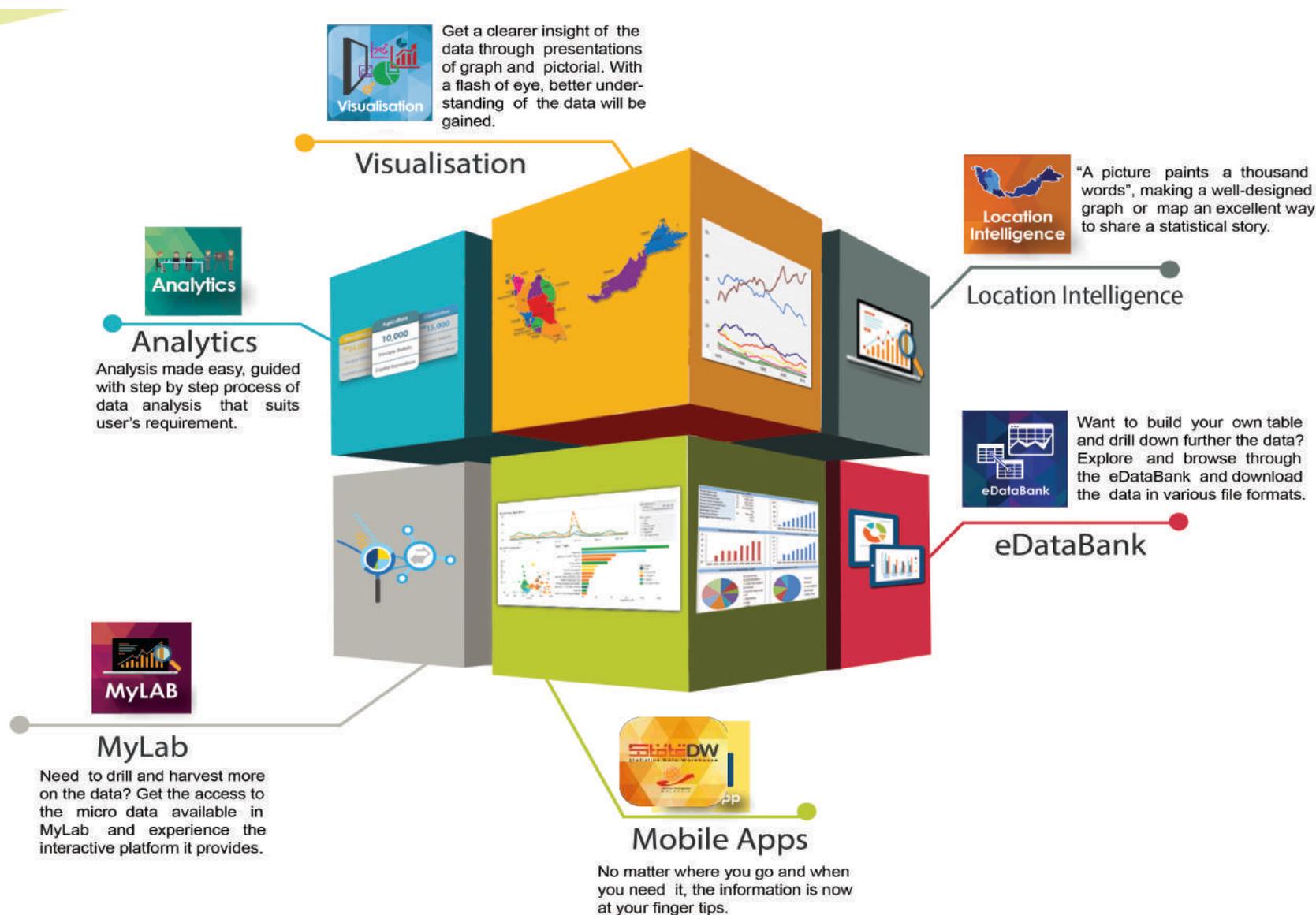


# ARCHITECTURE





# MODULES StatsDW PHASE I





# StatsDW Phase II

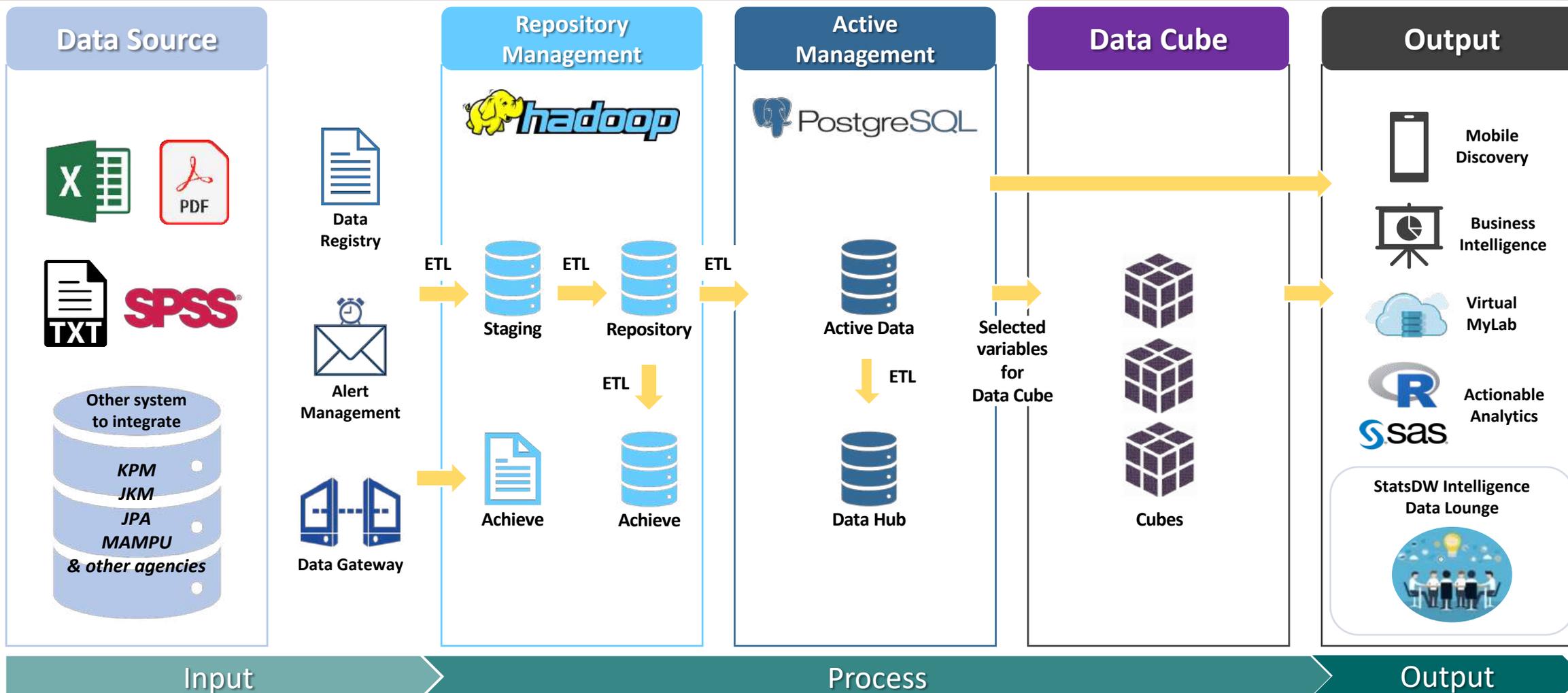


  
StatsMalaysia  
[www.DOSM.gov.my](http://www.DOSM.gov.my)





# StatsDW PHASE II ARCHITECTURE



**Note :**

KPM – Ministry of Education

JKM – Welfare Department

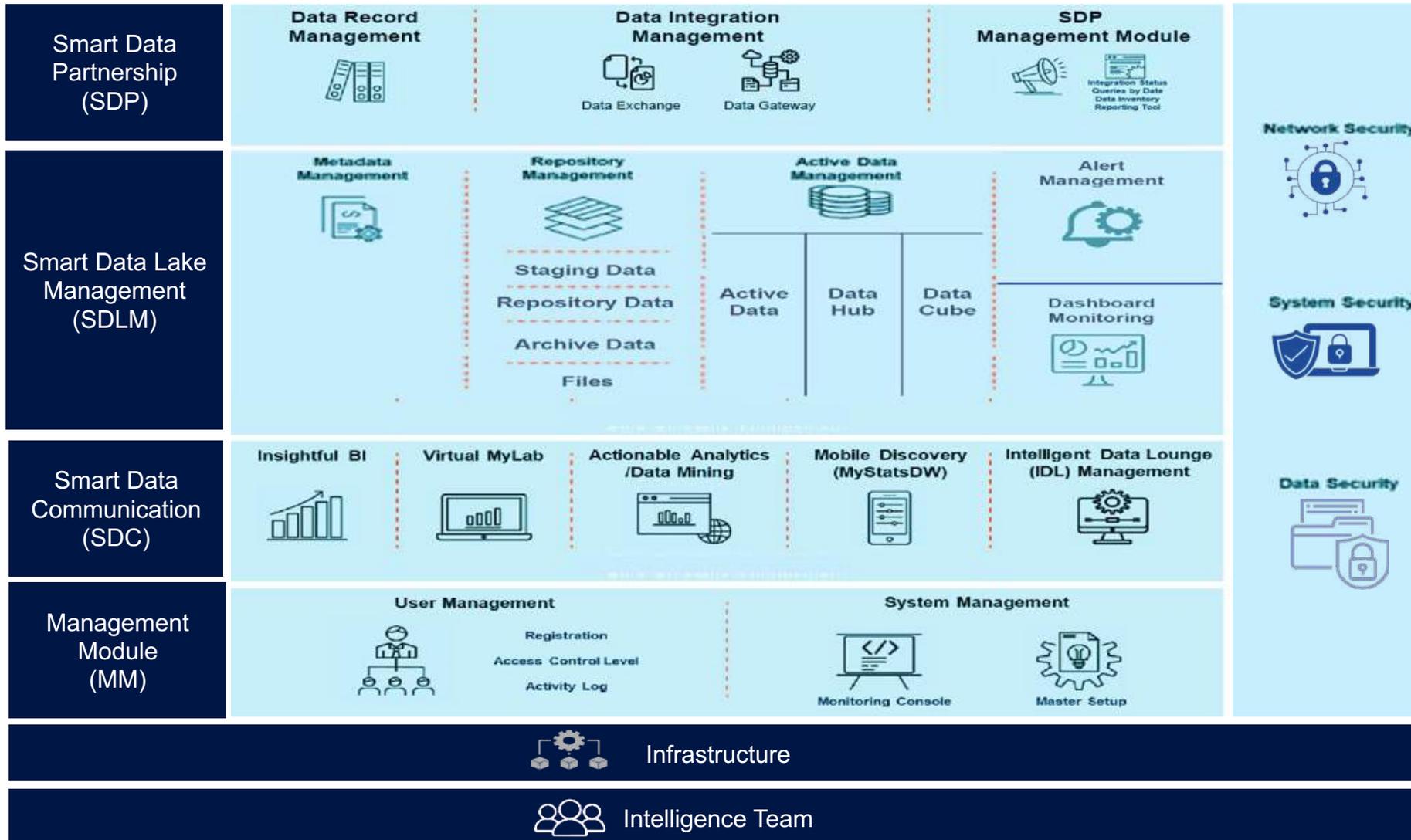
JPA – The Public Service Department

MAMPU – Malaysia Administrative Modernisation and Management Planning Unit

# StatsDW PHASE II MODULES

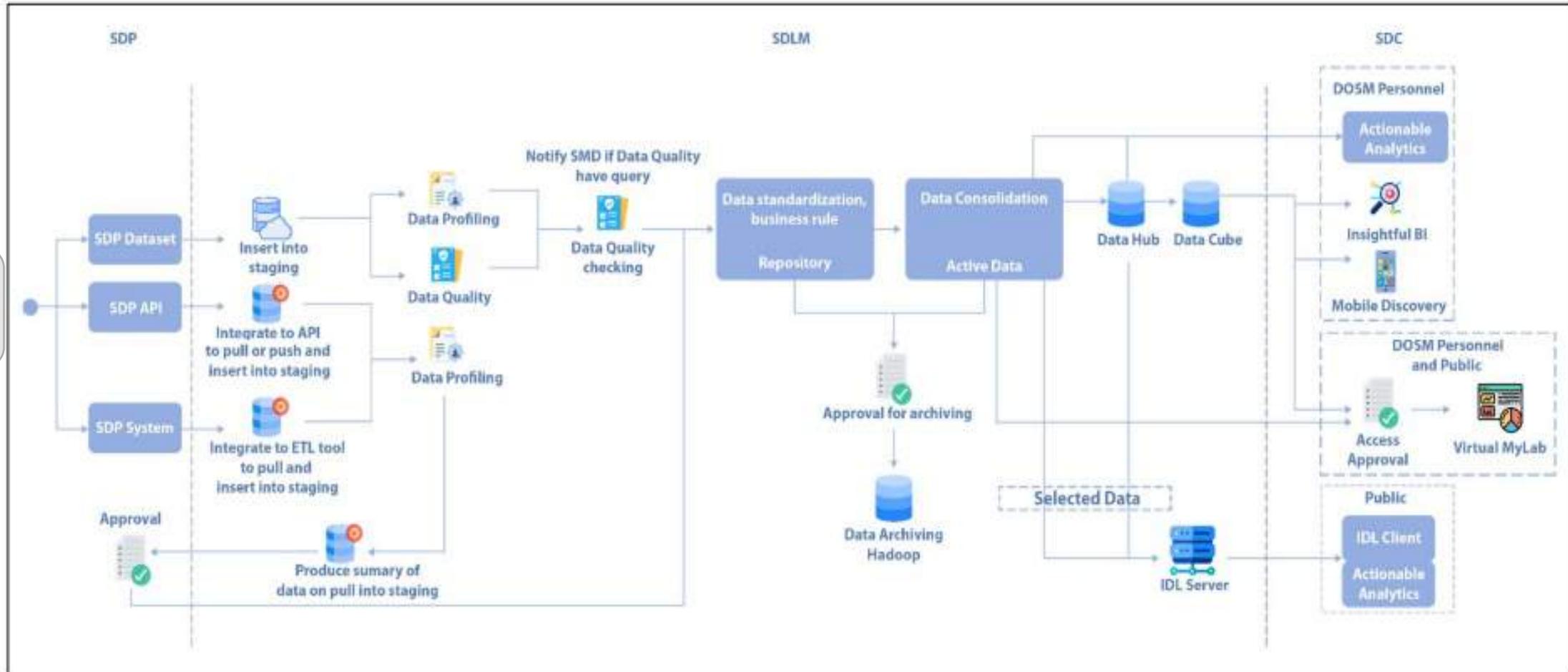


StatsDW Phase II will facilitate the storing, retrieving, processing activities of data for more efficient organising, administration, governing and securing of datasets





# DATA FLOW DIAGRAM



- Direct API
1. BNM
  2. MET M'sia
  3. JKM

# SMART DATA PARTNERSHIP (SDP)



## SDP

consists of 3 sub-modules which are:



### Data Record Management

Users able to register, submit & edit for acquisition by StatsDW 2.0 environment



### Data Integration Management

Integration of data with internal DOSM systems & other agencies via API. Users able to download data from repository & staging environment



### SDP Management Module

Dashboard for notification alerts and alerts & monitoring of data movement process from acquisition to storage



## User's Perspective

### Data Record Management

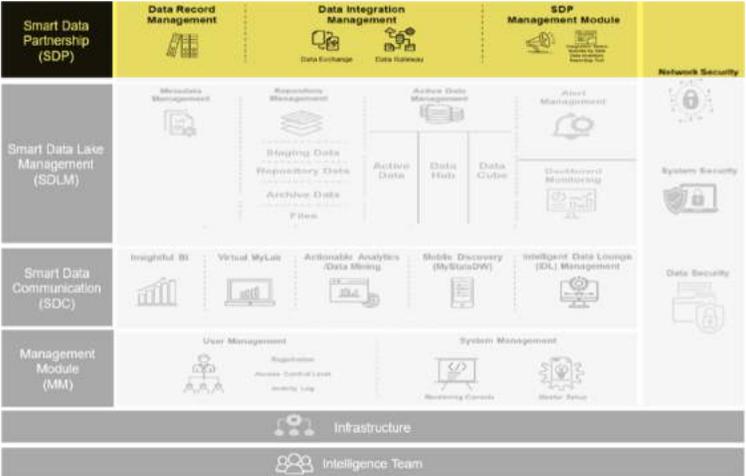
- 1** Dataset registration  
Includes metadata, technical notes, questionnaire
- 2** Director's role to **approve dataset** to be registered
- 3** **Dissemination level** for the registered dataset – to be decided by the SMD  
Whether the registered dataset to be stored in repository only or to be both, stored and visualized

### Data Exchange Management

- 1** Dataset uploading  
Includes metadata, technical notes, questionnaire (structured format)
- 2** If the dataset registered in the DR is not tally with the uploaded dataset, SMD will be notified via Alert Management

### Data Gateway Management

- 1** **StatsDW®** integration with other systems – in DOSM and OA
- 2** Director's role to **approve the integration** that will be conducted





# DATA RECORD MANAGEMENT



## Monitoring Dashboard



## Dataset registration

### HOUSEHOLD EXPENDITURE INFO

GENERAL

Data ID: v.1.000000 Registered By: provider  
 Agency: DOSM Division Approved By:  
 Division: BPHPP  
 Submitted Date: 13/05/2022  
 Submitted Time: 15:17

DATASET INFO

Dataset Name: Household Expenditure Status: Final  
 Data Source: DOSM Data Type: Micro  
 Source Type: Survey Theme: Social  
 Source Name: Household Expenditure Survey Report Sector: Household Expenditure  
 Dataset Frequency: Periodically Sub-Sector: -  
 Period: Yearly Action: Update  
 Reference Date: 2019 Dissemination Level: Full

## Record Layout Mapping

### HOUSEHOLD EXPENDITURE INFO

RECORD LAYOUT CONTROL FIGURE

Description  Data type  Code  Remarks

Matched  Unmatched  All

ACTIVE DATA ACCESSIBILITY

Variable Name		Description		Code		Unmatch reason
Previous	Current	Previous	Current	Previous	Current	
HID			Wujudkan anonymous HID number			
Relasiwang			Perhubungan Dengan Kelas (R-Rumah 01 - Kelas 01)			
Gender			JAKSA 1 - Laki; 2 - Perempuan			
Age			Umur			
KBT				Matched: 110		
Ethnic			Kumpulan etnik 1 - Melayu; 2 - Cina; 3 - India; 4 - Lain			
Marital_status			Taraf Perkahwinan 1 - Tidak pernah berkahwin; 2 - Berkahwin			
Citizenship			Kewarganegaraan 1 - Warganegara; 3 - Bukan Warganegara			



## SDLM Module Functions



Able to store structured and unstructured data, as well as files with Hadoop equipped - an open source large storage system, and complemented with high-storage capacity.



Stores and enables user control of data throughout its lifecycle in the system.



Facilitate to store, retrieve, process activities of data by using Extract Transform & Load (ETL) tools.



Deliver better solutions in preparing data, providing alerts and monitoring data activities.



Monitoring dashboard of all data processes and every data element is given a unique identifier and tagged with metadata.




## User's Perspective

### Metadata Management

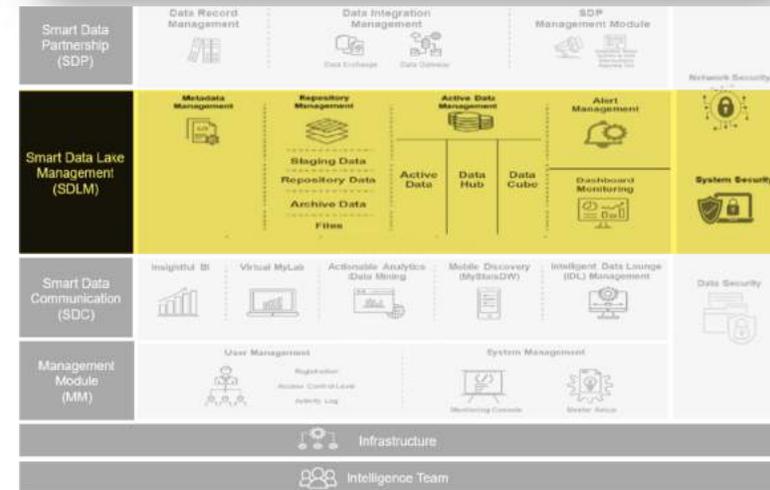
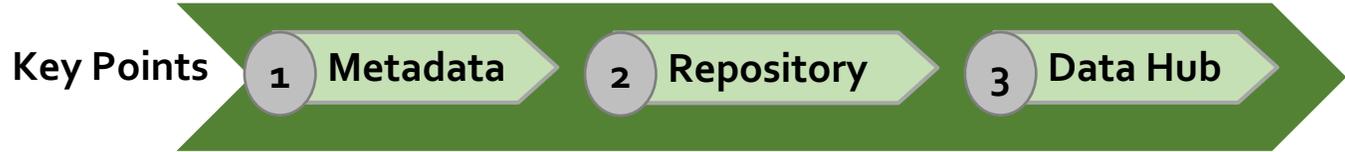
- 1 Metadata input to be key in by SMD  
If query – SMD to amend
- 2 Only query free metadata will be uploaded

### Repository Management

- 1 SMD has to amend the data structure of the dataset if there is query from the RM
- 3 All datasets that is uploaded by SMD will be stored in RM

### Active Data Management

- 1 SMD may request their desired hub in the ADM  
Data hub will be created based on user's request





# EXTRACT TRANSFORM LOAD (ETL) PROCESS



## Storing

### Load Table Processes

← Back

+ NEW PROCESS



Search

Name	Created At	
1) Load Source file to Hadoop Staging	12/16/2021 6:27:52 PM	
1.1) Load Source Member file to Hadoop Staging	1/10/2022 11:08:44 AM	
2) Transform table to suggested column name	12/22/2021 3:12:20 PM	
4) Lookup Table	2/26/2022 11:53:28 AM	

### Flow Processes

Back

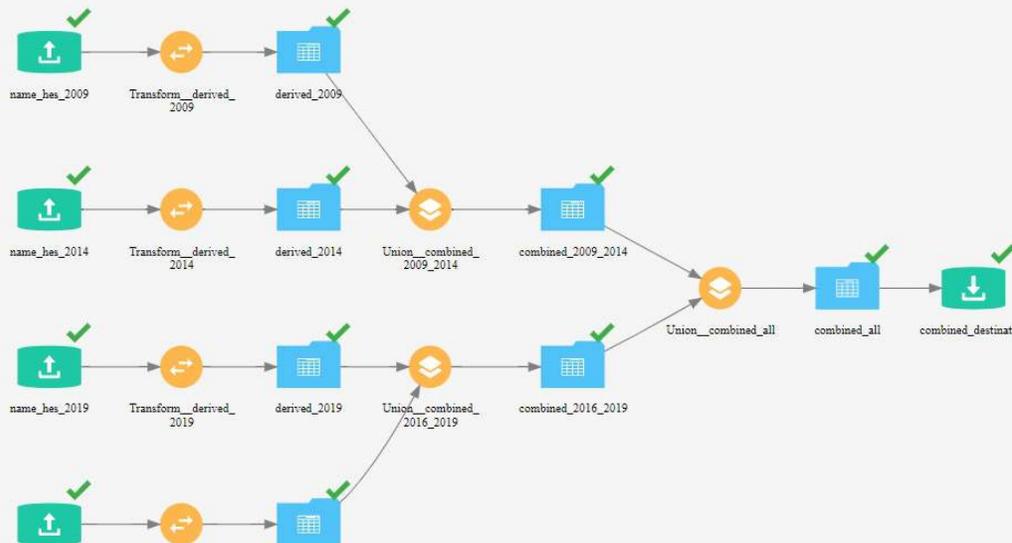
+ NEW PROCESS



Name

- 3) Combine Original Source with Member data
- 5) Combine Year →
- 5) Combine Year (Latest)
- 6) Data Cleansing and Data Elements
- 7) Repository
- 8) Active Data
- 9) Datahub

### 5) Combine year - Household Expenditure Survey



## Processing

-7 of 7

Navigation icons: back, forward, search, etc.



# EXTRACT TRANSFORM LOAD (ETL) PROCESS



## Household\_Expenditure Survey

### Load source table

Column mappings

Source: dsT022019HES  
Destination: ds\_bphpp\_hes\_2019\_q

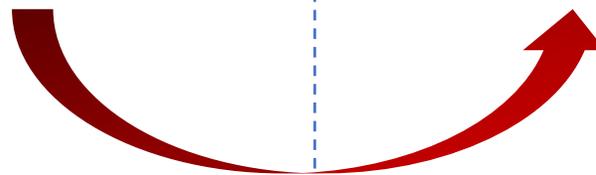
Source	Destination	Source Type
<input checked="" type="checkbox"/>	HID	TEXT
<input checked="" type="checkbox"/>	Relationship	TEXT
<input checked="" type="checkbox"/>	Gender	TEXT
<input checked="" type="checkbox"/>	Age	DOUBLE
<input checked="" type="checkbox"/>	KET	TEXT

### Transform table to suggested column name

Column mappings

Source: ds\_bphpp\_hes\_2009\_combine  
Destination: name\_household\_expenditure\_2009  
Metadata: Metadata\_HES\_2009

Source	Destination	Source Type	Destination Type
<input checked="" type="checkbox"/>	hid	STRING	string
<input checked="" type="checkbox"/>	relationship	DOUBLE	double
<input checked="" type="checkbox"/>	gender	DOUBLE	double
<input checked="" type="checkbox"/>	age	DOUBLE	double
<input checked="" type="checkbox"/>	ket	DOUBLE	double





# EXTRACT TRANSFORM LOAD (ETL) PROCESS



## Data standardization & Business rule

Transform\_\_cleansing

← BACK

SAVE

A unik_id	A perhubungan_c	A jantina	# umur	A kumpulan_etnik	A grp_kumpulan	A taraf_perkahwii	# kewarganegara	A persekolahan	A taraf_pendidika	A sijil_tertinggi_d	A taraf_aktiviti	A penerima_pend	A pekerjaan	# pekerjaan_mas	A indu
K5150201010300	1.0	1.0	57	1.0	1.0	2.0	1	3.0	4.0	9.0	4.0	1.0	61321.0	6	11
S089120712024C	1.0	1.0	37	1.0	1.0	2.0	1	3.0	8.0	3.0	2.0	1.0	51499.0	5	75
J0210102540130	1.0	1.0	82	2.0	2.0	2.0	1	1.0	13.0	8.0	6.0	1.0	003.0	10	00
D063030210670C	1.0	1.0	56	1.0	1.0	2.0	1	1.0	13.0	8.0	3.0	1.0	91211.0	9	95
C094060407040C	1.0	1.0	56	1.0	1.0	2.0	1	3.0	6.0	1.0	9.0	1.0	004.0	10	00
J1200102070150C	1.0	1.0	73	1.0	1.0	2.0	1	3.0	3.0	9.0	4.0	1.0	61322.0	6	11
M0470402150070C	1.0	1.0	31	1.0	1.0	2.0	1	3.0	8.0	3.0	4.0	1.0	91423.0	9	45
S0601207390340C	1.0	2.0	21	8.0	1.0	1.0	1	3.0	8.0	3.0	3.0	1.0	42123.0	4	52
N0609504060030C	1.0	2.0	81	1.0	1.0	3.0	1	3.0	2.0	9.0	6.0	1.0	003.0	10	00
S2541207120710C	1.0	1.0	28	12.0	4.0	1.0	2	3.0	5.0	9.0	3.0	1.0	72307.0	7	50
Q543130720000C	1.0	1.0	41	1.0	1.0	2.0	1	3.0	4.0	9.0	3.0	1.0	51442.0	5	75
M09304010000C	1.0	1.0	53	1.0	2.0	2.0	1	3.0	8.0	3.0	3.0	1.0	52201.0	5	52

Transformation Scripts

```

load
settype col:perhubungan_dengan_ketua_isi_rumah
type:'integer'
settype col:perhubungan_dengan_ketua_isi_rumah
type:'string'
settype col:jantina type:'integer'
settype col:jantina type:'string'
settype col:kumpulan_etnik type:'integer'
settype col:kumpulan_etnik type:'string'
set col:kumpulan_etnik value:'0' row:
empty(kumpulan_etnik)
settype col:grp_kumpulan_etnik_utama type:'integer'
settype col:grp_kumpulan_etnik_utama type:'string'
settype col:taraf_perkahwinan type:'integer'
settype col:taraf_perkahwinan type:'string'
settype col:persekolahan type:'integer'
settype col:persekolahan type:'string'
settype col:taraf_pendidikan_nasmi_tertinggi type:'integer'
settype col:taraf_pendidikan_nasmi_tertinggi type:'string'
settype col:sijil_tertinggi_diperolehi type:'integer'
settype col:sijil_tertinggi_diperolehi type:'string'
settype col:taraf_aktiviti type:'integer'
settype col:taraf_aktiviti type:'string'
settype col:penerima_pendapatan type:'integer'
settype col:penerima_pendapatan type:'string'
settype col:pekerjaan type:'integer'
settype col:pekerjaan type:'string'
set col:pekerjaan value:'0' row:empty(pekerjaan)
settype col:industri type:'integer'

```

### Data standardization :

The user shall use the master data file, which located in the repository database for data standardization. Data standardization is the process of converting data to a common format to enable users to process and analyze it.

i.e: Standardize and group data by age range

scl_age_main_id	scl_age_main_desc_e	scl_age_main_desc_m	CASE
1	0-14	0-14	WHEN UMUR BETWEEN 0 AND 14 THEN 1
2	15-64	15-64	WHEN UMUR BETWEEN 15 AND 64 THEN 2
3	65 and over	65 ke atas	WHEN UMUR BETWEEN 65 AND 134 THEN 3
			END

### Business Rule:

A business rule is a statement that imposes some form of constraint on a specific aspect of the database, such as the elements within a field specification for a particular field or the characteristics of a given relationship.

i.e :

- i. Household Income Survey: starting in 2007, 30% of the number of households with 22 variables.
- ii. Population and Housing Census: starting in 1980, 2% of the total population data of Malaysia.



## User's Perspective

### Insightful BI

- 1 SMD have to determine the dashboard requirement
  - IBI admin will develop the dashboard
  - SMD will do the dashboard checking in the development server
- 2 If SMD add new variable which is not in the data hub yet, they have to go through DR process
- 3 External dashboard and Internal Dashboard

### Mobile Discovery

- 1 User can do lite analysis
- 2 User can drag and drop variables, and dashboard can be visualized in mobile view
- 3 Can be accessed (free), using Android and IOS

### Virtual MyLab

- 1 Single Sign On (SSO) access through DOSM Active Directory
- 2 User can download the output

Key Points

1 Dashboard

2 Virtual

3 Mobile

# SMART DATA COMMUNICATION (SDC)



## User's Perspective

### Actionable Analytic



- 1 Can be access by internal (DOSM environment) and external (IDL)
- 2 User can search, drag and drop the required data/survey/census/ variable
- 3 The chosen data can be put in the data cart
- 4 Statistical software tools will be provided for the user to do the analysis
- 5 Generated script can be stored by the user

### IDL

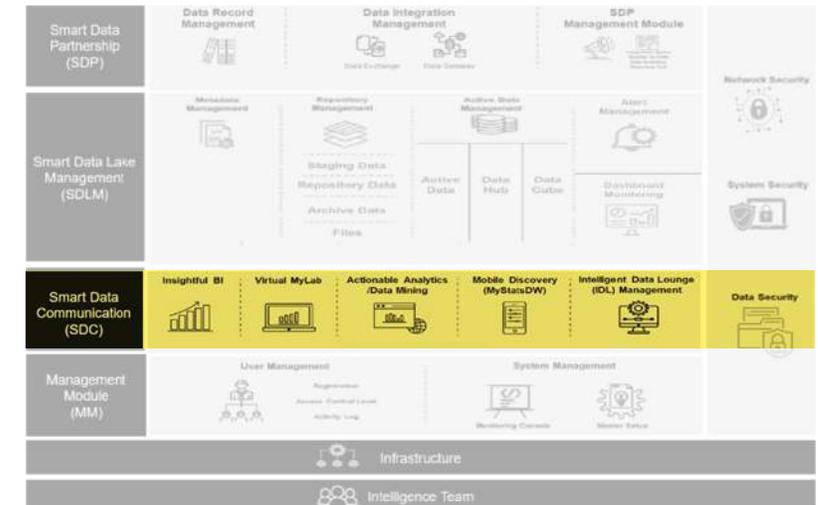


- 1 Cyber cafe concept
- 2 User can register online for booking or walk in
- 3 No time limit for IDL usage (during office hours)



### Key Points

- 1 Data cart
- 2 IDL





## Various Avenues Users Can View And/Or Perform Analytics:

### 01. INSIGHTFUL BI

Through a web platform @ portal, information will be delivered through Business Intelligence (BI) form. Users (internally and externally), can view reports, predefined dashboards and download reports based on user-provided variables

### 02. VIRTUAL MYLAB

Selected data for statistical purposes are accessible through this avenue. It comes with an interactive table builder feature for registered users, to use with BI tools. Since this module is restricted to only internal DOSM users, Access Control Level (ACL) is applied, and unauthorised users must request permission. Users can create reports, dashboards and transform data into interactive presentations. Internal users can view metadata and perform analysis

### 03. ACTIONABLE ANALYTICS (AA)

Selected micro data are included here for analytical/data mining for authorised users. ACL is implemented, to limit access to external parties. Authorised users can view metadata, perform analysis, to script and to code. **AA** is accessible via the Intelligent Data Lounge (IDL), subject to DOSM's micro data dissemination policy



### 06. STATSDW PORTAL

Main platform to access statistical data, reports and the dashboard. This portal contains four services; Actionable Analytics, Virtual MyLab, Insightful BI and Mobile Discovery and can be accessed by the public

### 05. INTELLIGENT DATA LOUNGE (IDL) MANAGEMENT

Provides on-site access to Actionable Analytics, Virtual MyLab and Insightful BI to external users to experience

### 04. MOBILE DISCOVERY (MYSTATSDW)

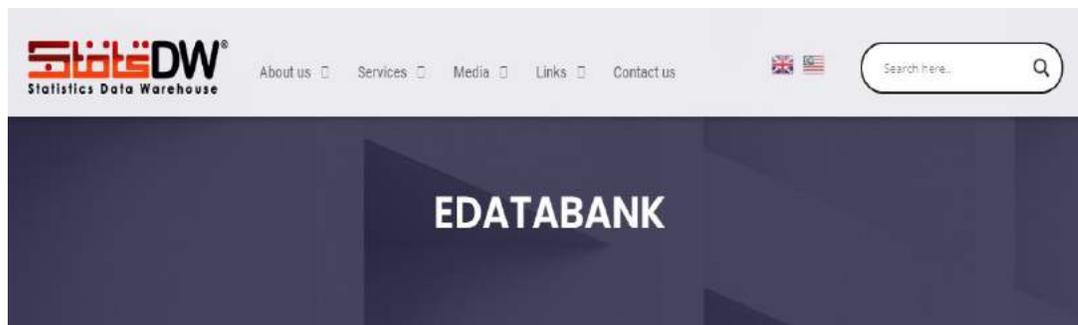
Statistics are delivered in an interactive graphical form that is tailored for mobile devices for both internal and external users. Visualisation outputs are similar to Insightful BI



## Main page Insightful BI

The screenshot shows the main page of the StatsDW (Statistics Data Warehouse) website. The header includes the StatsDW logo, navigation links (About us, Services, Media, Links, Contact us), a search bar, and flags for the United Kingdom and Malaysia. The main content area is titled 'SERVICES' and features a grid of service tiles:

- DATA CATALOG** (Icon: Document with magnifying glass)
- EDATABANK** (Icon: Bank building)
- DATA VISUALISATION** (Icon: Monitor with line graph)
- MOBILE DISCOVERY** (Icon: Smartphone with signal waves)
- VIRTUAL MYLAB** (Icon: Computer monitor with data points)
- ACTIONABLE ANALYTICS** (Icon: Head with gears)
- INTELLIGENCE DATA LOUNGE** (Icon: Camera and document)



[Home](#) [Services](#) [eDataBank](#)

## EDATABANK

eDataBank is an electronic data bank of Department of Statistics Malaysia (DOSM). Users may compile their own tables and graphs and export these tables in different file formats. The information can be accessed, downloaded and printed.

[BPAN](#)

[BMP](#)

[BPE](#)

[BPIP](#)

[BPPA](#)

[BPPAS](#)



[About us](#) [Services](#) [Media](#) [Links](#) [Contact us](#)



Search here..

## EDATABANK

eDataBank is an electronic data bank of Department of Statistics Malaysia (DOSM). Users may compile their own tables and graphs and export these tables in different file formats. The information can be accessed, downloaded and printed.

[BPAN](#)

Consist of 10 Main List to choose, all related reports located in the main list also included.

### 1. Annual National Accounts

1. Annual GDP at Current Prices by Kind of Economic Activity (RM Million, Annual percentage change, Percentage share)
2. Annual GDP at Constant Prices by Kind of Economic Activity (RM Million, Annual percentage change, Percentage share)
3. Annual GDP at Current Prices by Type of Expenditure (RM Million, Annual percentage change, Percentage share)
4. Annual GDP at Constant Prices by Type of Expenditure (RM Million, Annual percentage change, Percentage share)
5. GDP by Kind of Economic Activity at Current Prices – RM Million

### 2. Quarterly National Accounts

1. Quarterly GDP at Current Prices by Kind of Economic Activity (RM Million, Percentage change from corresponding quarter of preceding year, Percentage change from preceding quarter, Percentage share)
2. Quarterly GDP at Constant Prices by Kind of Economic Activity (RM Million, Percentage change from corresponding quarter of preceding year, Percentage change from preceding quarter, Percentage share)
3. Quarterly GDP at Current Prices by Type of Expenditure (RM Million, Percentage change from corresponding quarter of preceding year, Percentage change from preceding quarter, Percentage share)
4. Quarterly GDP at Constant Prices by Type of Expenditure (RM Million, Percentage change from corresponding quarter of preceding year, Percentage change from preceding quarter, Percentage share)

### 3. State Socioeconomic Report

1. GDP by State at Constant Prices (RM Million, Annual Percentage Change and Percentage Share)
2. GDP by State and Kind of Economic Activity at Constant Prices (RM Million, Annual Percentage Change and Percentage Share)
3. GDP by State at Current Prices (RM Million)
4. GDP per Capita by State at Current Prices (RM)



## EDATABANK – BPAN : ANNUAL NATIONAL ACCOUNTS

### EDATABANK – BPAN : ANNUAL NATIONAL ACCOUNTS

eDataBank is an electronic data bank of Department of Statistics Malaysia (DOSM). Users may compile their own tables and graphs and export these tables in different file formats. The information can be accessed, downloaded and printed.

[eDataBank – BPAN : Annual National Accounts](#)

[eDataBank – BPAN : Quarterly National Accounts](#)

[eDataBank – BPAN : State Socioeconomic Report](#)

[eDataBank – BPAN : Information & Communication Technology Satellite Account](#)

BPAN : Annual National Accounts

Annual National Accounts GDP by Kind of Economic Act... ▾

1. Year :

(All)

2020

2019

2018

2. Pick the Kind of Economic Activity :

(All)

Agriculture

Construction

Manufacturing

Mining and quarrying

GDP by Kind of Economic Activity at Current Prices - RM Million

Kind Of Economic Activity	RM Million								
	2008	2012	2013	2014	2015	2016	2017	2019	2020
Agriculture	76,753	95,122	92,830	98,177	97,539	105,756	117,995	109,549	116,084
Construction	21,156	37,909	42,692	48,650	55,382	61,089	66,552	71,067	57,537
Manufacturing	189,105	224,730	232,659	253,087	262,379	272,396	299,797	323,862	315,608
Mining and quarrying	120,170	101,474	103,484	109,064	103,059	103,959	121,326	131,117	95,328
Services	355,233	501,830	536,186	585,518	643,883	689,969	747,360	860,497	816,292



# INSIGHTFUL BI - DATA VISUALIZATION



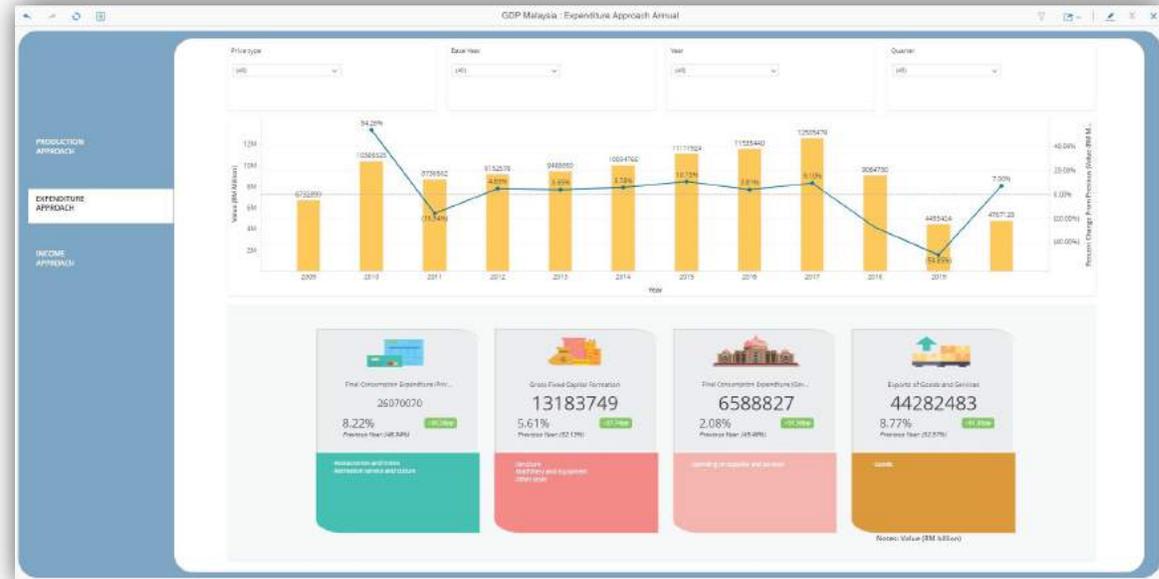
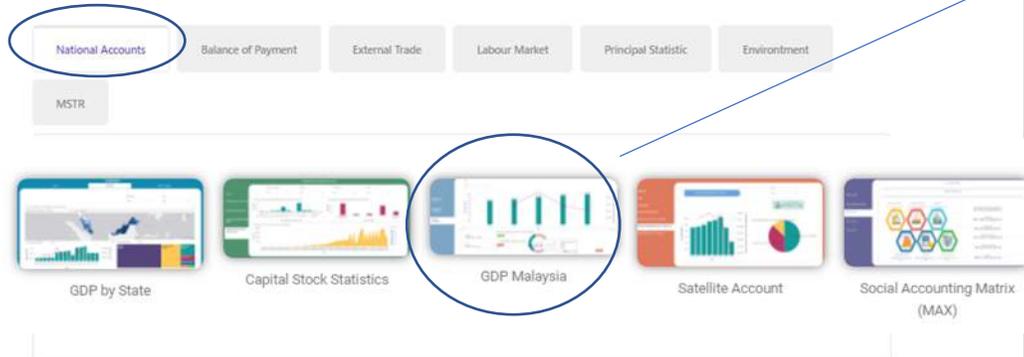
## Data Visualisation – National Accounts



Home > Services > Data Visualization

### DATA VISUALIZATION

Department of Statistics Malaysia has traditionally presented its data in many ways: in articles, press releases, tables, databank, maps, bar charts, pie charts, and line graphs. New interactive visualisation techniques were first introduced in 2015, with the Industrial Production Index and the Population Pyramid as the best known examples. These visualisations give a clearer insight into how important data on a specific area interact.

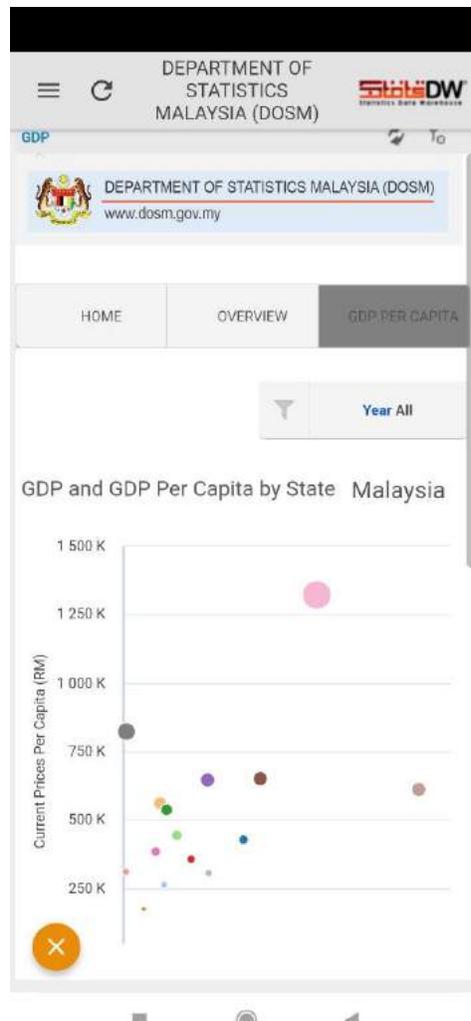
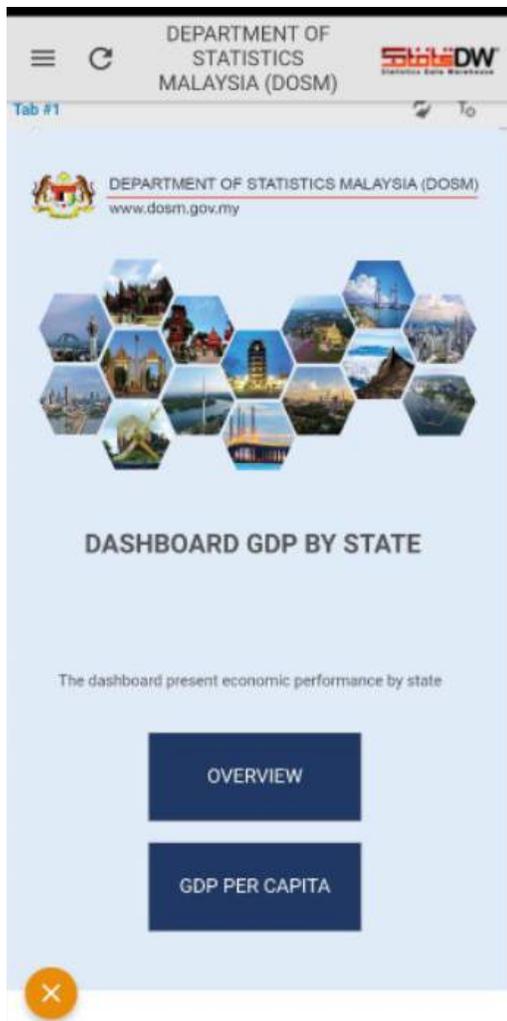




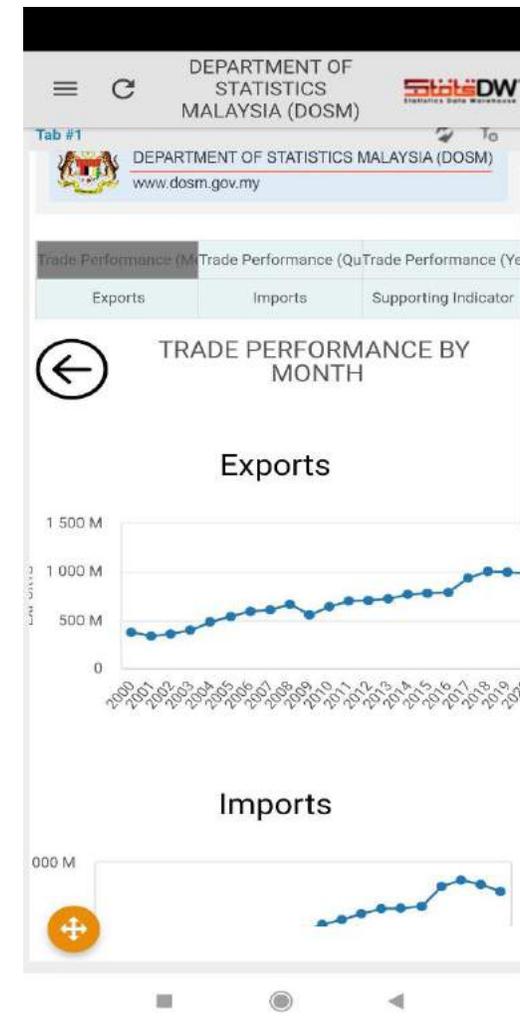
# MOBILE DISCOVERY



## GDP



## EXTERNAL TRADE





# VIRTUAL MYLAB

PRIME MINISTER'S DEPARTMENT  
DEPARTMENT OF STATISTICS MALAYSIA



StatsMalaysia  
www.DOSM.gov.my



**Virtual MyLab Page**

Virtual MyLab Requests

Request Variable - have unchecked requests. <span>New</span>	May 12 2022
1 Request Variables	DELETE
Request Variable - have unchecked requests. <span>New</span>	May 11 2022
1 Request Variables	DELETE
Request Variable - have unchecked requests. <span>New</span>	May 18 2022
1 Request Variables	DELETE

**1. Click 'Go To Virtual My Lab'**

MyLab > Shared Reports > H\_MBLS

**Create**

**Labour\_Salaries\_MBLS**  
Owner: Administrator  
Modified: 4/22/22 11:42:06 AM

**New Dossier**  
Owner: Administrator  
Modified: 3/14/22 8:29:22 AM

Recents  
Shared Reports  
H\_MBLS  
My Reports  
My Objects  
MicroStrategy Library

**3. User will get a credential of MicroStrategy to get access of the data.**

**Statistics Data Warehouse**

**VIRTUAL MYLAB**  
Virtual mylab is web-based platform where users use data in the catalogue to create the dashboard and report using Microstrategy Dossier features.

**Subject**  
Economy  
Social  
Labour Market  
Social  
new subject  
New Subject 1  
hghghf

**Topic**  
Demography  
Gender Statistics  
Household Expenditure  
Environment Statistics  
Social Statistics  
Migration  
Population

**Variable**  
 Origin Strata  
 Ethnic  
 Citizenship  
 Marital Status  
 Occupation  
 Industry  
 Education Level  
 Summary  
 Employment Level

**2. User can choose Subject, Topic and Variable. Then, proceed to request variable.**

SEND REQUEST CANCEL

**Visualization 1**

Jumlah Gaji Dan Upah Tersembunyi (Kerja Sambilan)

Tahun	Jumlah Gaji Dan Upah Tersembunyi (Kerja Sambilan)
2010	~900
2011	~1000
2012	~1050
2013	~1100
2014	~1150
2015	~1200
2016	~1150
2017	~1450
2018	~1550
2019	~1750
2020	~1650

**4. User can create their own Visualisation and share it.**



# ACTIONABLE ANALYTICS

PRIME MINISTER'S DEPARTMENT  
DEPARTMENT OF STATISTICS MALAYSIA

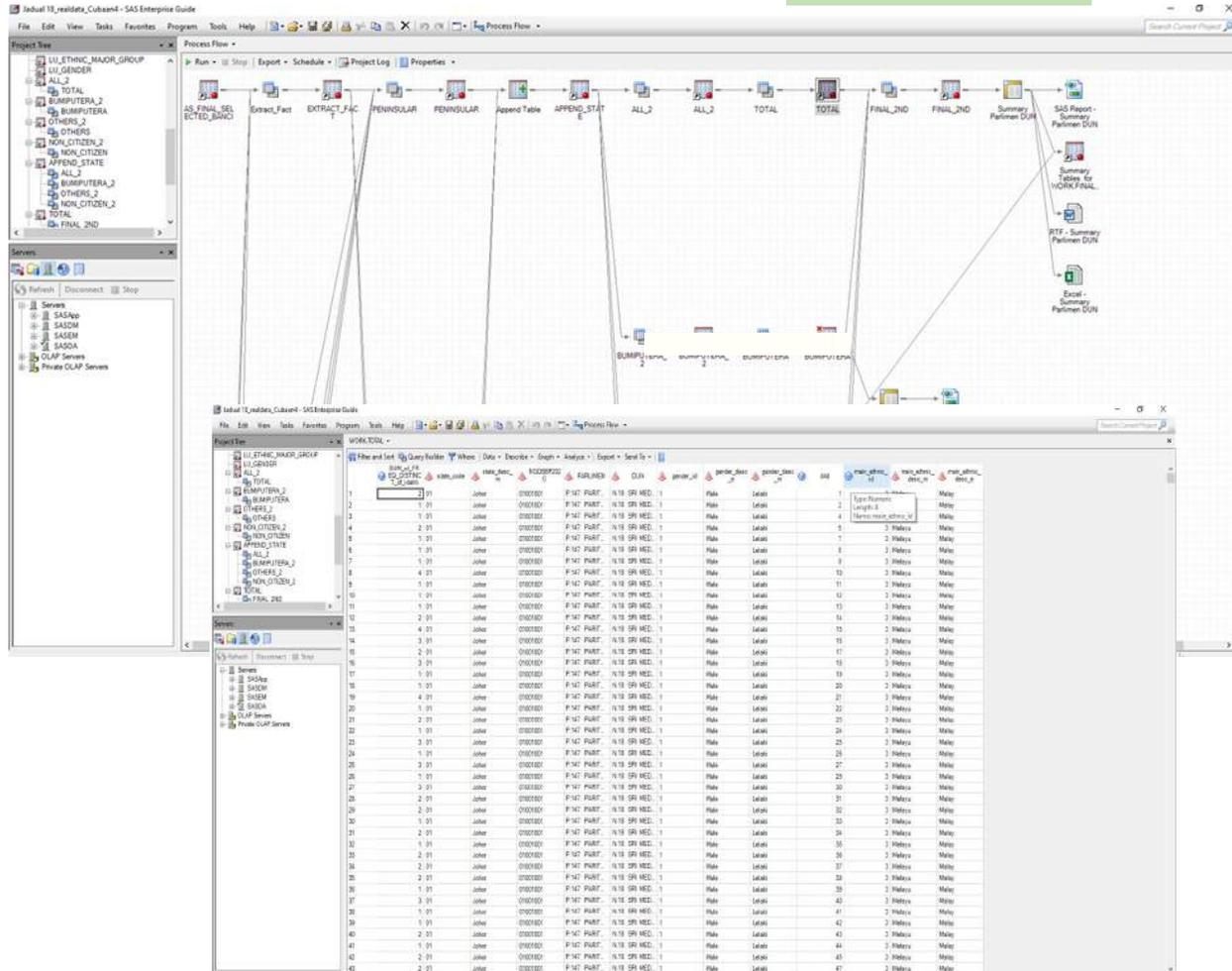


     
StatsMalaysia  
www.DOSM.gov.my



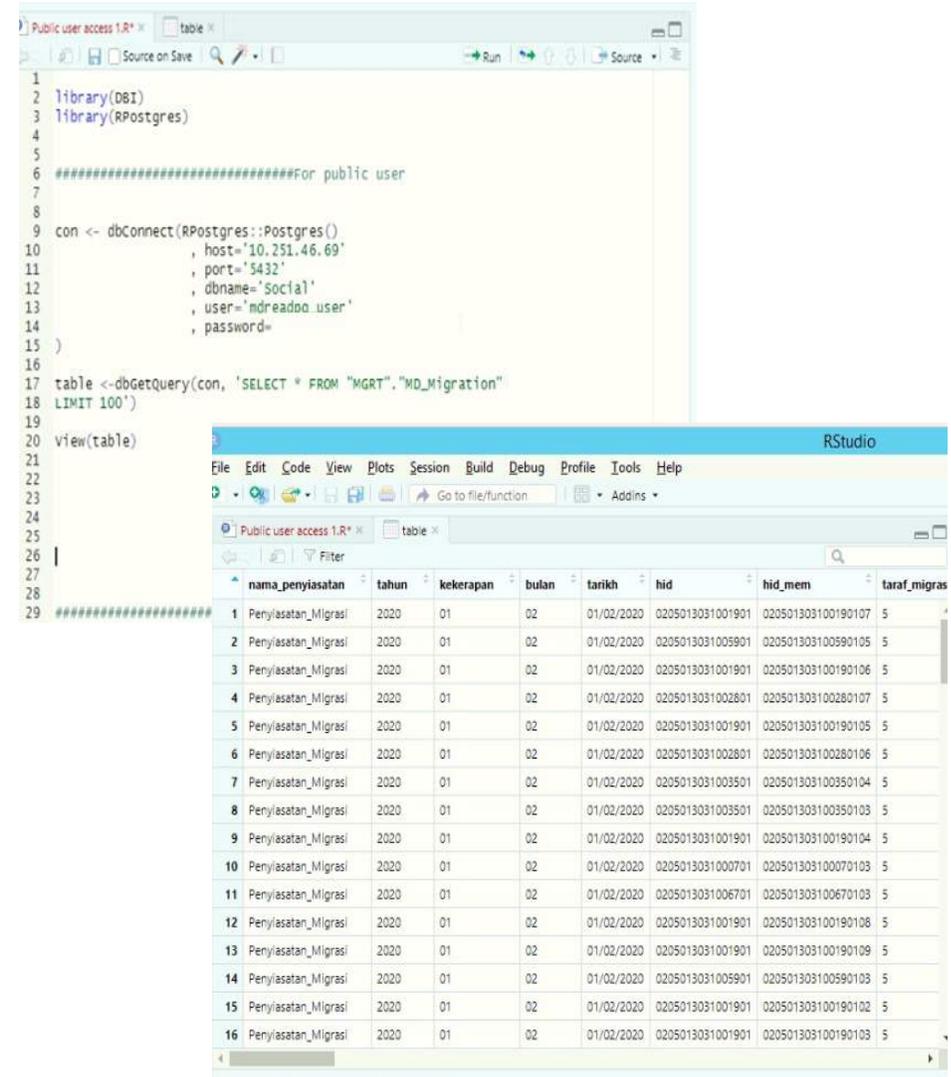
Actionable Analytics is a platform where users request data in the catalogue to conduct analysis using statistical tools, either R Studio or SAS EG

## SAS EG



The screenshot displays the SAS Enterprise Guide (EG) interface. The top window shows a 'Project Flow' diagram with steps: AS\_FINAL\_SELECTED\_BANG, Extract\_Fact, EXTRACT\_FACT, PENINSULAR, PENINSULAR, Papered Table, APPEND\_STATE, ALL\_2, ALL\_2, TOTAL, TOTAL, FINAL\_ZND, FINAL\_ZND, Summary Parlimen DUN, Summary Parlimen DUN, SAS Report - Summary Parlimen DUN, Summary Tables for WORK\_FINAL, RTF - Summary Parlimen DUN, and Excel - Summary Parlimen DUN. The bottom window shows a 'WORK TABLE' with columns: L1\_ETHNIC\_MAJOR\_GRP, L1\_GENDER, ALL\_2, BUMPUPERA\_2, BUMPUPERA\_2, OTHERS\_2, OTHERS\_2, NON\_CITIZEN\_2, NON\_CITIZEN\_2, ALL\_2, BUMPUPERA\_2, BUMPUPERA\_2, OTHERS\_2, OTHERS\_2, NON\_CITIZEN\_2, NON\_CITIZEN\_2, and FINAL\_ZND. The table contains 42 rows of data.

## R Studio



The screenshot displays the R Studio interface. The top window shows R code for connecting to a PostgreSQL database and querying a table. The bottom window shows the resulting data table.

```

1
2 library(DBI)
3 library(RPostgres)
4
5
6 #####For public user
7
8
9 con <- dbConnect(RPostgres::Postgres(
10   , host='10.251.46.69'
11   , port='5432'
12   , dbname='social'
13   , user='mdreadoo_user'
14   , password=
15 )
16
17 table <- dbGetQuery(con, 'SELECT * FROM "MGRT"."MD_Migration"
18 LIMIT 100')
19
20 view(table)
21
22
23
24
25
26
27
28
29 #####

```

nama_penyasatan	tahun	kekerapan	bulan	tarikh	hid	hid_mem	taraf_migras
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031001901	020501303100190107	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031005901	020501303100590105	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031001901	020501303100190106	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031002801	020501303100280107	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031001901	020501303100190105	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031002801	020501303100280106	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031003501	020501303100350104	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031003501	020501303100350103	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031001901	020501303100190104	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031000701	020501303100070103	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031006701	020501303100670103	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031001901	020501303100190109	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031000701	020501303100070103	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031005901	020501303100590103	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031001901	020501303100190102	5
Penyasatan_Migrasi	2020	01	02	01/02/2020	0205013031001901	020501303100190103	5

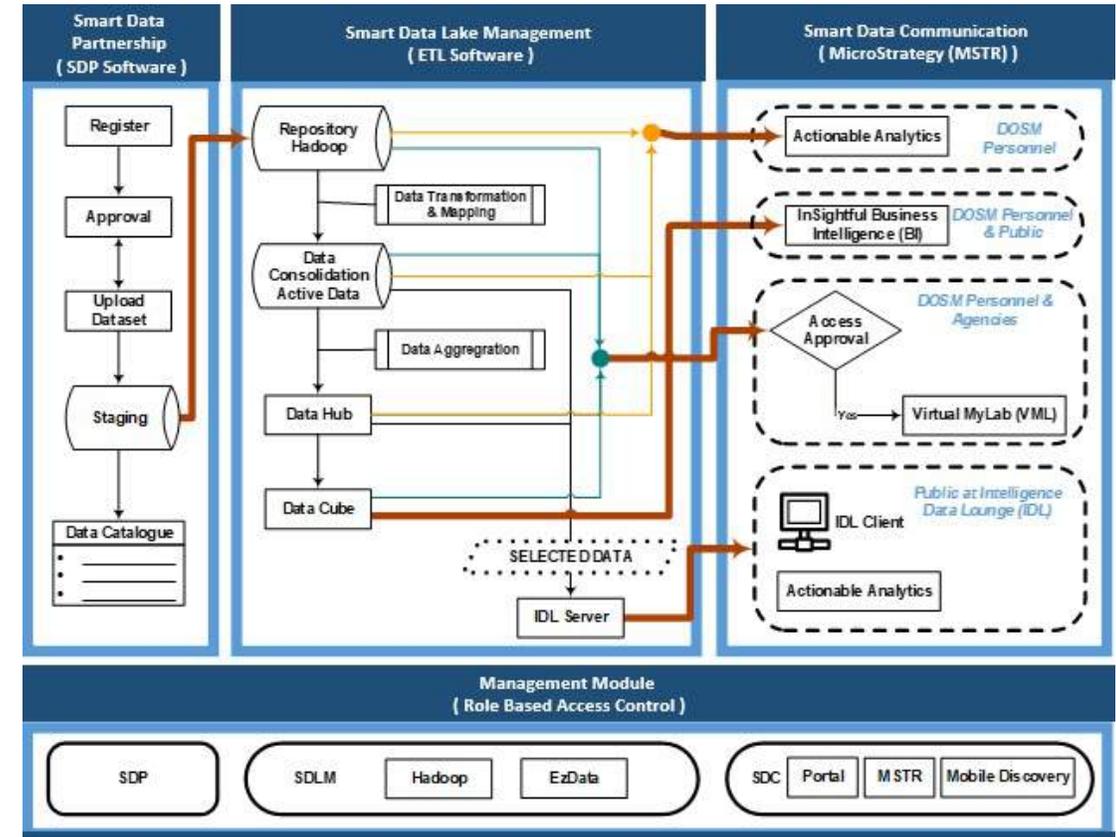
# MANAGEMENT MODULE (MM)



A platform to centralized administration of the StatsDW Phase II:

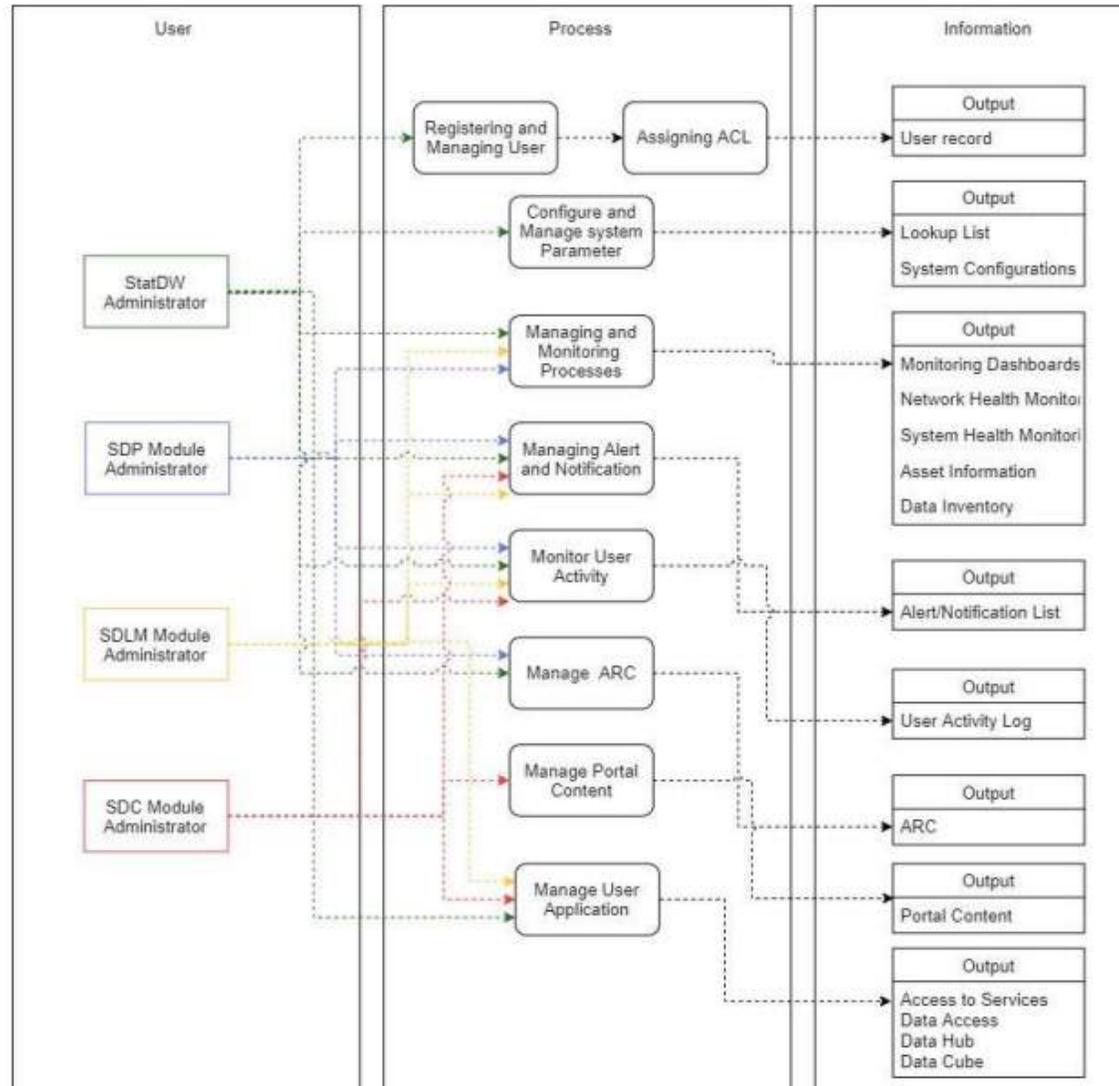
MM allows normal business process of StatsDW Phase II to be conducted fully on **system and user management**.

- ❖ **SDP** consists of the customized platform to monitor the whole process of storing the data recorded with alerts and notifications.
- ❖ **SDLM** consists of a Hadoop framework as a database involving massive amounts of data and computation that is scalable to store structured, unstructured data and files; and EzData as the data management and ETL tool.
- ❖ **SDC** consists of Portal, MSTR to perform analytics BI system which will use Active Data as a source for cubes; and Mobile Discovery as a mobile application delivering statistics in an interactive graphical form.



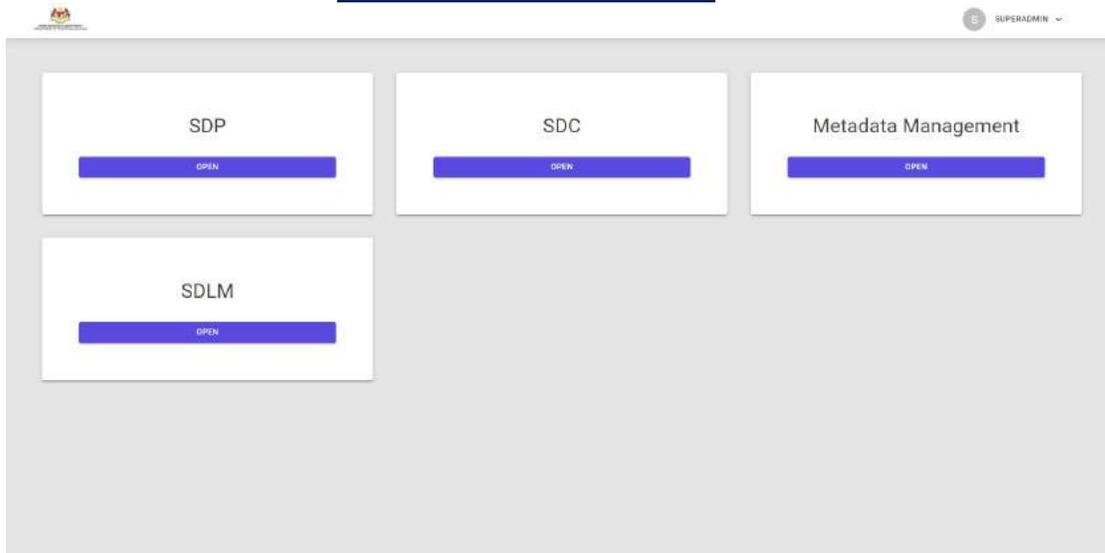


# INFORMATION ARCHITECTURE





## System Admin

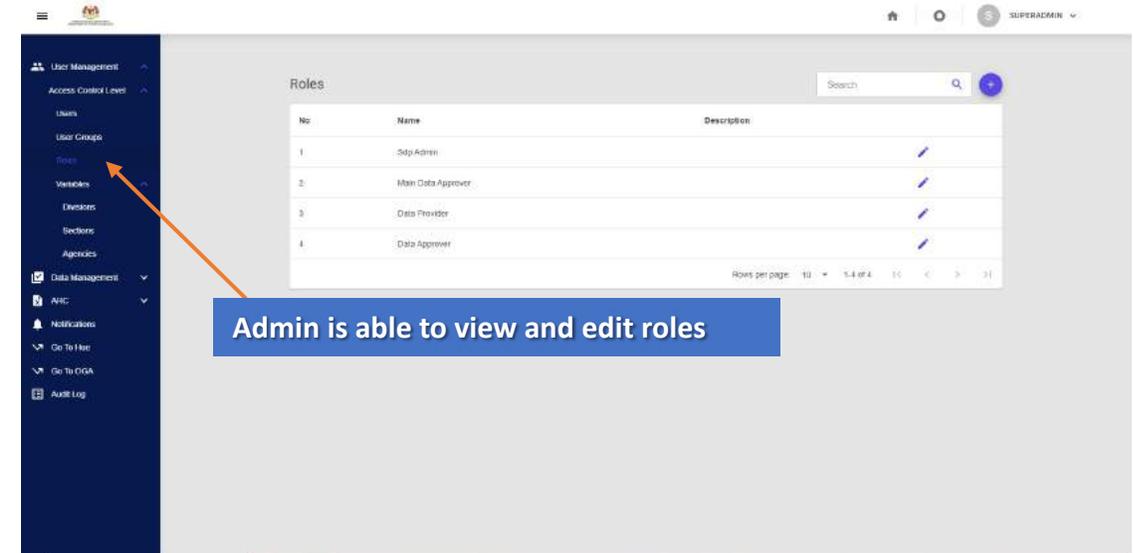


SDP OPEN

SDC OPEN

Metadata Management OPEN

SDLM OPEN

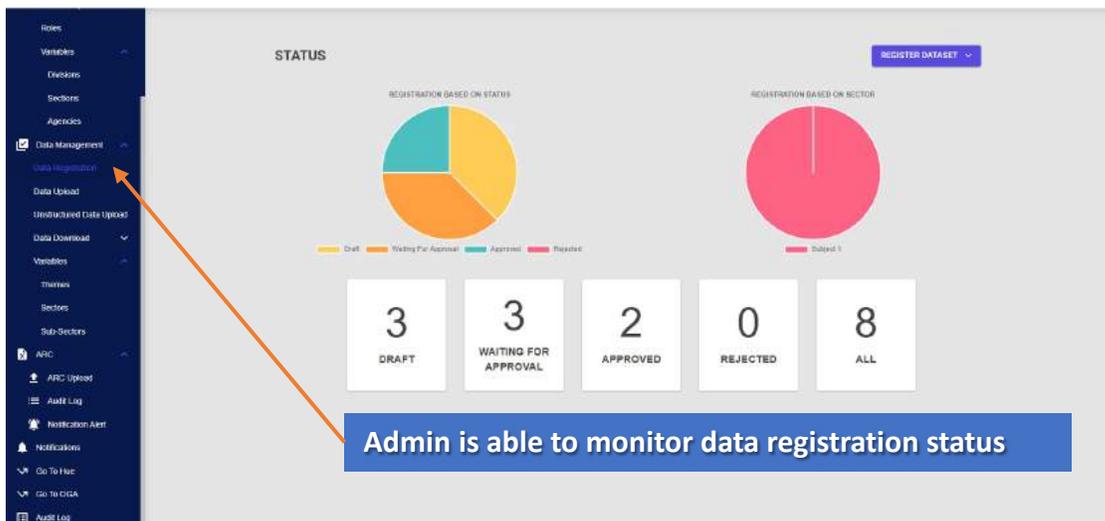


Roles

No	Name	Description
1	Sdp Admin	
2	Mlan Data Approver	
3	Data Provider	
4	Data Approver	

Rows per page: 10 1-4 of 4 10 < > 31

Admin is able to view and edit roles



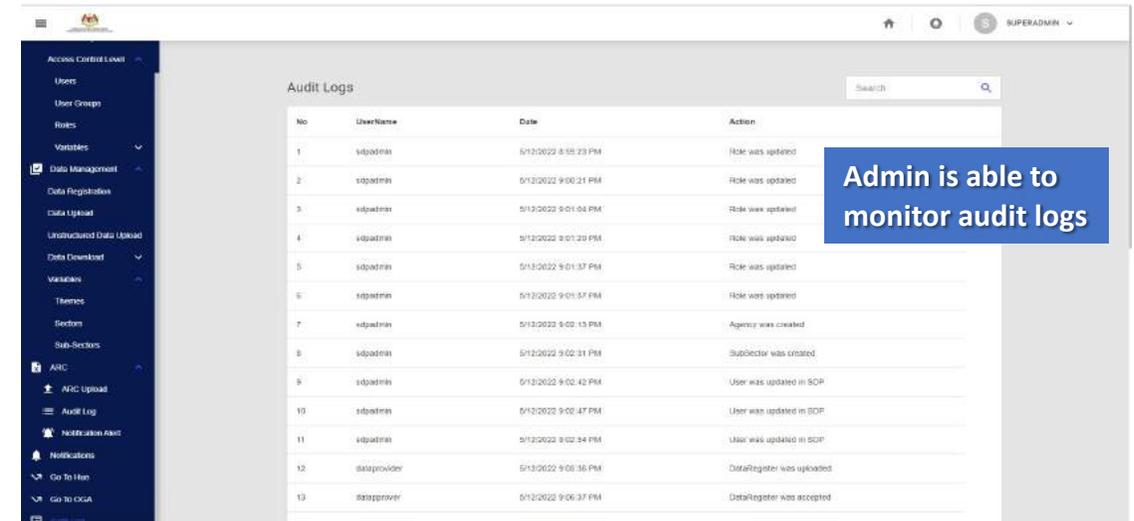
STATUS REGISTER DATASET

REGISTRATION BASED ON STATUS

REGISTRATION BASED ON SECTOR

3 DRAFT 3 WAITING FOR APPROVAL 2 APPROVED 0 REJECTED 8 ALL

Admin is able to monitor data registration status



Audit Logs

No	Username	Date	Action
1	sdpadmin	01/12/2022 8:58:23 PM	Role was updated
2	sdpadmin	01/12/2022 9:00:21 PM	Role was updated
3	sdpadmin	01/12/2022 9:01:04 PM	Role was updated
4	sdpadmin	01/12/2022 9:01:20 PM	Role was updated
5	sdpadmin	01/12/2022 9:01:37 PM	Role was updated
6	sdpadmin	01/12/2022 9:01:57 PM	Role was updated
7	sdpadmin	01/12/2022 9:02:13 PM	Agency was created
8	sdpadmin	01/12/2022 9:02:31 PM	Subsector was created
9	sdpadmin	01/12/2022 9:02:42 PM	User was updated in SDP
10	sdpadmin	01/12/2022 9:02:47 PM	User was updated in SDP
11	sdpadmin	01/12/2022 9:02:54 PM	User was updated in SDP
12	sdaprovider	01/12/2022 9:03:36 PM	DataRegister was uploaded
13	sdaprovider	01/12/2022 9:06:37 PM	DataRegister was accepted
14	malmetapprover	01/12/2022 9:07:05 PM	DataRegister was accepted

Admin is able to monitor audit logs



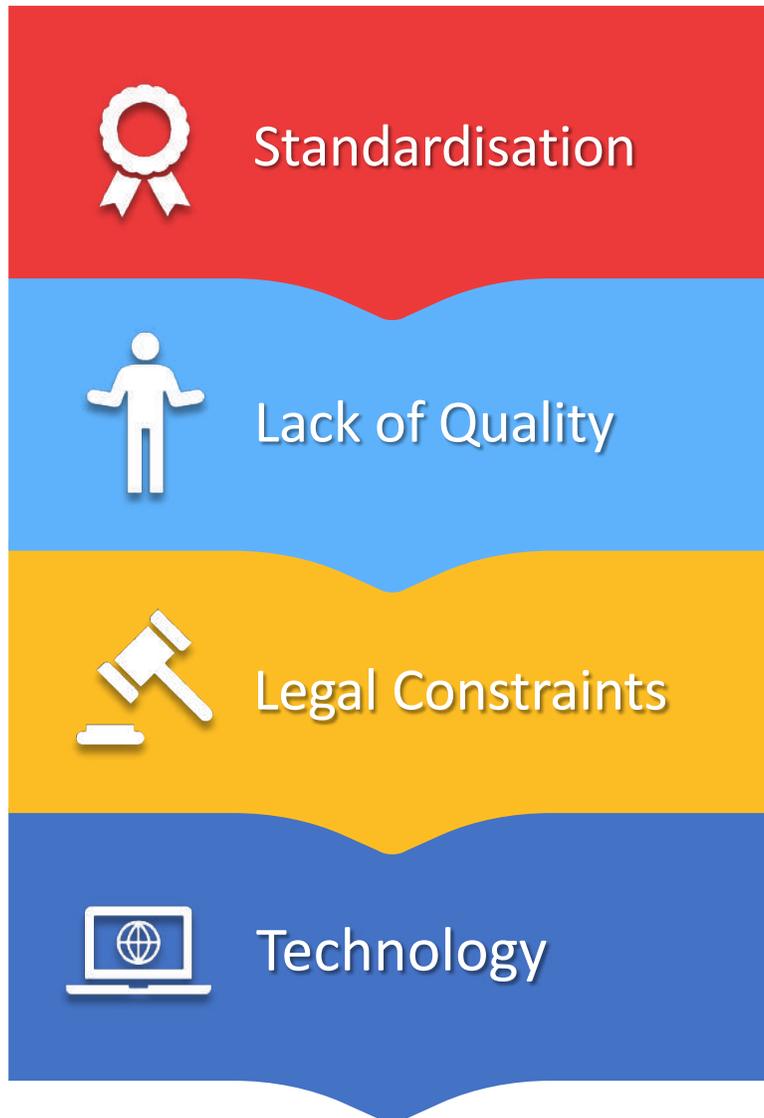
# CHALLENGES



     
StatsMalaysia  
[www.DOSM.gov.my](http://www.DOSM.gov.my)



 #KELUARGA  
MALAYSIA



## Standardisation

- Data in various format
- Changes of data structure

## Lack of Quality

- Data inconsistency

## Legal constraints

- Data sharing on individual data

## Technology capabilities

- Cost is on the rise
- Technology capabilities



# INNOVATION IMPACT



  
StatsMalaysia  
[www.DOSM.gov.my](http://www.DOSM.gov.my)



 #KELUARGA  
MALAYSIA



# INNOVATION IMPACT

PRIME MINISTER'S DEPARTMENT  
DEPARTMENT OF STATISTICS MALAYSIA



     
StatsMalaysia  
www.DOSM.gov.my





# INNOVATION IMPACT



1  
Displaying data of quality, integrity and assurance of its validity

2  
Platform of reference for DOSM

3  
Simplify data generation and analysis

4  
Preserving the confidentiality of individual data

5  
Publish interactive table and visualisation

6  
Save time and cost of data search





# BENEFITS AND WAY FORWARD OF StatsDW





# BENEFITS OF StatsDW



## BENEFIT

- Provide centralized data storage centre
- Maintaining data reliability and quality
- Facilitate wider access and analysis
- Guarantees data security
- Integrate systems and data sharing between agencies
- Improved data integration, management and security
- Improved metadata management
- Facilitates development of new analytical outputs and products
- Facilitates National Statistical Service objectives



**StatsDW**<sup>®</sup>  
Statistics Data Warehouse





# BENEFITS OF StatsDW



## Data Driven Society



**IMPLEMENTATION OF PUBLIC SECTOR OPEN DATA (Pekeliling AM Bil. 1/2015)**

- Creation of new industry or innovation through the digital economy.
- Transparency in service delivery.
- Better service delivery





- Improvement in bottom –up Data Mart to maximize Subject Matter Divisions ( in DOSM) the capability and capacity to create own Data Mart
- Data Governance and Master Data Repository: a central repository for other government agencies in Malaysia
- Data Homogenization: Similarity of data format, classification and metadata among agencies
- Extension of Statistical Data and Metadata Exchange (SDMX): Moving towards the new standard of SDMX 3.0

# THANK YOU



StatsMalaysia

[www.DOSM.gov.my](http://www.DOSM.gov.my)



2022



2012-2022



20 OKT



2016-2030

