StatsDW

Statistics Data Warehouse

Towards Streamlining The Data Management

By:

Nur Hurriyatul Huda binti Abdullah Sani

27-28 August 2015

INSTITUT LATIHAN STATISTIK MALAYSIA

DEPARTMENT OF STATISTICS MALAYSIA

CONTENT

1. Introduction
2. StatsDW Objective
3. Issues & Challenges
4. Way Forward

Department of Statistics Malaysia as Producer of National Statistics Agency

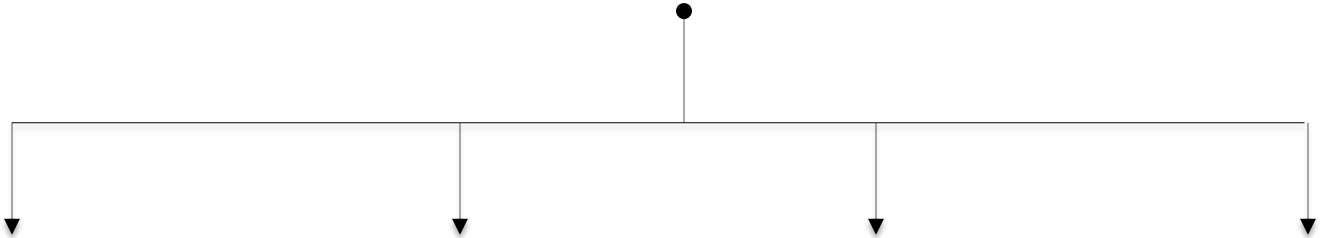
Function

•Under the "Statistics Act 1965 (Revised 1989)", the main functions of the department are: To collect and interpret statistics for the purpose of formulation or implementation of government policies in whatever fields as needed by the government or for fulfilling the requirements of trade, commerce, industry, agriculture or others

•To disseminate statistics which have been collected or interpretation based on statistics collected, not only to government agencies but also to authorities or persons where the information is useful to them

INTRODUCTION

SITUATION IN DOSM



Data Management System in DOSM built and saved by the application and Subject.

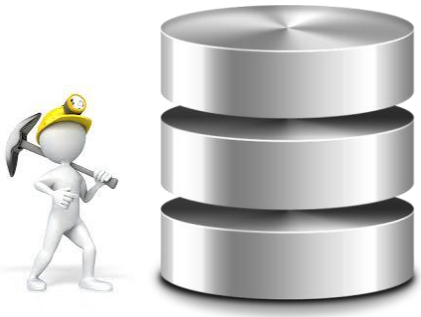
Data is stored in a silo

Historical data (micro and aggregate) are stored separately in the Subject Matter Division respectively

The issue of data publication:
a. softcopy / hardcopy missing
b. stored in various locations.

INTRODUCTION

IMPLICATION

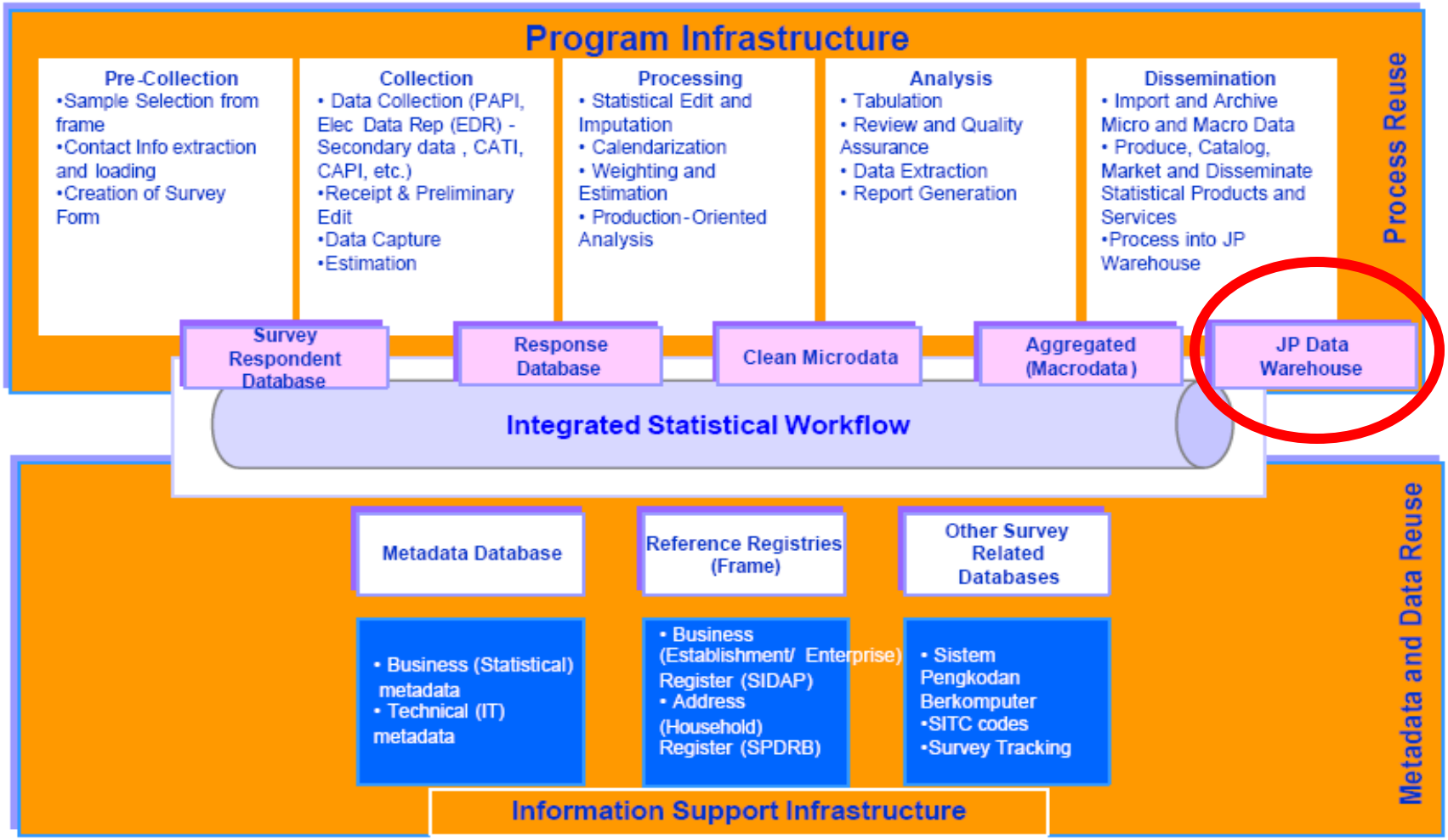


Limited data mining / harvesting

Less efficient data management

Data archiving is not systematic

INTRODUCTION



StatsDW OBJECTIVES

1. To consolidate all Historical Micro Data and Aggregated Data in the Enterprise Data Warehouse
2. To enhance data quality and consistency
3. To enable fast and easy access to data stored
4. To disseminate the data to wider users in a timely manner

ISSUE & CHALLENGES

ISSUE & CHALLENGES

DATA MANAGEMENT

- Data Source
- Data Structure / Record Layout
- System / Platform
- Generation Issue
- Where & Which the Final Data ?

CAPACITY ABILITY

- System Integration & Interaction
- Data Preparation Duration
- Data Access
- Storage Capacity

TECHNOLOGY

- Data warehouse tools
- Migration tools
- Business Intelligence
- User Acceptance

USER ACCESS

- Access Control Level

SECURITY

- ICT Security Policy
- Statistic Act 1965 (Revised 1989)
- Census Act 1960
- Micro Data Dissemination Policy

ISSUE & CHALLENGES

DATA MANAGEMENT

- Data Source
- Data Structure / Record Layout
- System / Platform
- Generation Issue
- Where & Which the Final Data ?

CAPACITY ABILITY

- System Integration & Interaction
- Data Preparation Duration
- Data Access
- Storage Capacity

TECHNOLOGY

- Data warehouse tools
- Migration tools
- Business Intelligence
- User Acceptance

SECURITY

- ICT Security Policy
- Statistic Act 1965 (Revised 1989)
- Census Act 1960
- Micro Data Dissemination Policy

ESS

- Access Cont

StatsDW ARCHITECTURE

Jabatan Perangkaan MALAYSIA

Data Sources

Structured Data

- Micro Data
- Aggregated Data
- Historical Data

Unstructured Data

Data Governance, Data Integration (ETL)

Data Profiling & Data Quality

Data Migration, Data Integration (ETL)

Enterprise Data Warehouse (EDW)

Data Warehouse Appliance

Operational Data Store

Module

Visualisation, eData Bank, Mobile

End Users

Internal User

External User

DATA MANAGEMENT

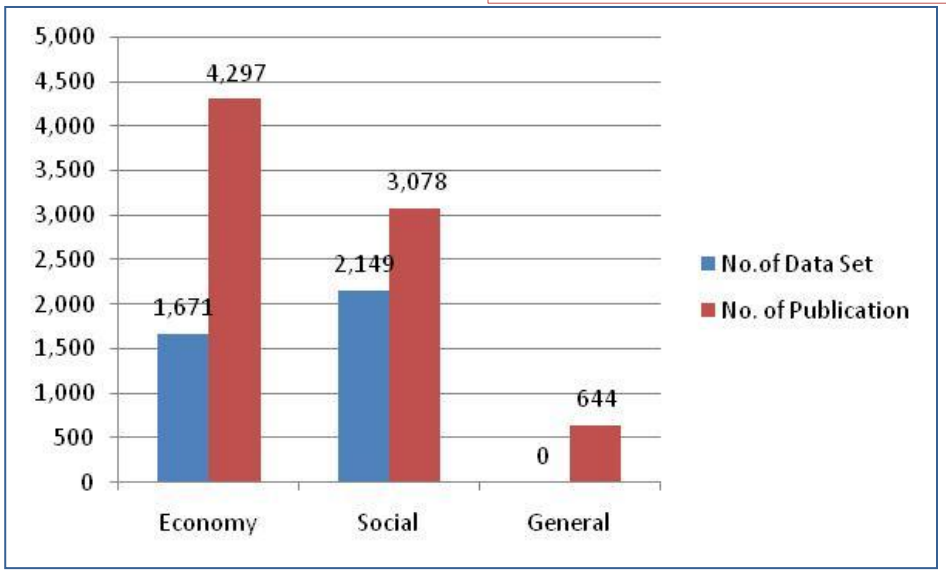
ISSUE & CHALLENGES

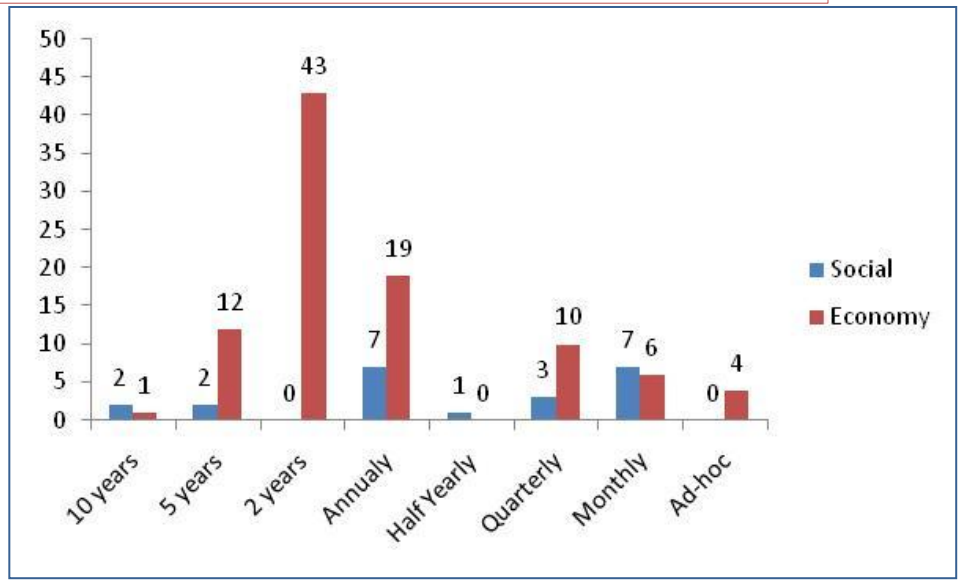
DATA MANAGEMENT

Data Source

Source diversification and data form :

- **Type of census/ survey ;**
- **Frequency of census/ survey ;**
- **Structured and unstructured data.**





Data Classification	No.of Data Set	No. of Publication
Economy	1,671	4,297
Social	2,149	3,078
General	0	644
Jumlah	**3,820**	**8,019**

Total number of data set and publication release by DOSM as at Jun 2015

ISSUE & CHALLENGES

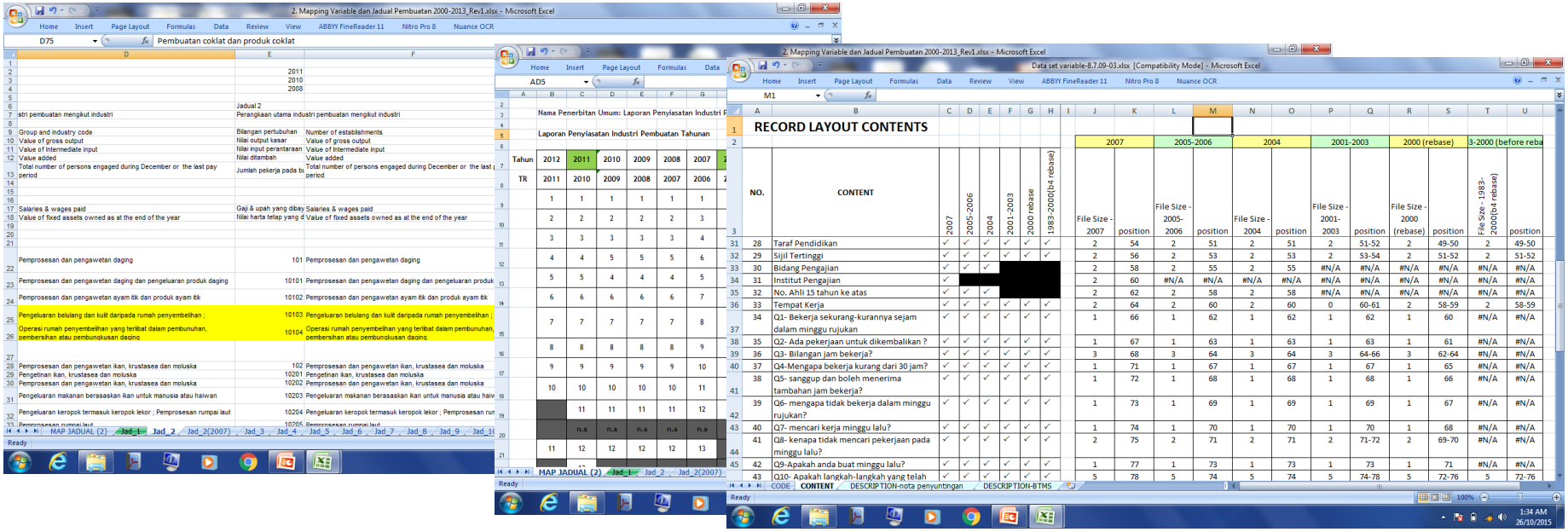
DATA MANAGEMENT

Data Structure / Record Layout

Changes found in:;

- **record layout and data structure;**
- **variables;**
- **questionnaire / survey form;**

Need to have data mapping for standard layout and structure.



2. Mapping Variable dan Jadual Pembuatan 2000-2013_Rev1.xlsx - Microsoft Excel

Pembuatan coklat dan produk coklat

	2011	
	2010	
	2008	
	Jadual 2	
stri pembuatan mengikut industri	Perangkaan utama industri pembuatan mengikut industri	
Group and industry code	Bilangan pertubuhan	Number of establishments
Value of gross output	Nilai output kasar	Value of gross output
Value of Intermediate input	Nilai input perantaraan	Value of Intermediate input
Value added	Nilai ditambah	Value added
Total number of persons engaged during December or the last pay period	Jumlah pekerja pada bu	Total number of persons engaged during December or the last period
Salaries & wages paid	Gaji & upah yang dibay	Salaries & wages paid
Value of fixed assets owned as at the end of the year	Nilai harta tetap yang d	Value of fixed assets owned as at the end of the year
Pemprosesan dan pengawetan daging	101	Pemprosesan dan pengawetan daging
Pemprosesan dan pengawetan daging dan pengeluaran produk daging	10101	Pemprosesan dan pengawetan daging dan pengeluaran produk
Pemprosesan dan pengawetan ayam itik dan produk ayam itik	10102	Pemprosesan dan pengawetan ayam itik dan produk ayam itik
Pengeluaran belulang dan kulit daripada rumah penyembelihan ;	10103	Pengeluaran belulang dan kulit daripada rumah penyembelihan ;
Operasi rumah penyembelihan yang terlibat dalam pembunuhan, pembershian atau pembungkusan daging	10104	Operasi rumah penyembelihan yang terlibat dalam pembunuhan, pembershian atau pembungkusan daging
Pemprosesan dan pengawetan ikan, krustasea dan moluska	102	Pemprosesan dan pengawetan ikan, krustasea dan moluska
Pengetinan ikan, krustasea dan moluska	10201	Pengetinan ikan, krustasea dan moluska
Pemprosesan dan pengawetan ikan, krustasea dan moluska	10202	Pemprosesan dan pengawetan ikan, krustasea dan moluska
Pengeluaran makanan berasaskan ikan untuk manusia atau haiwan	10203	Pengeluaran makanan berasaskan ikan untuk manusia atau haiw
Pengeluaran keropok termasuk keropok lekor ; Pemprosesan rumpai laut	10204	Pengeluaran keropok termasuk keropok lekor ; Pemprosesan run
Pemprosesan rumpai laut	10205	Pemprosesan rumpai laut

Nama Penerbitan Umum: Laporan Penyiasatan Industri P

Laporan Penyiasatan Industri Pembuatan Tahunan

Tahun	2012	2011	2010	2009	2008	2007	2
TR	2011	2010	2009	2008	2007	2006	2
	1	1	1	1	1	1	
	2	2	2	2	2	3	
	3	3	3	3	3	4	
	4	4	5	5	5	6	
	5	5	4	4	4	5	
	6	6	6	6	6	7	
	7	7	7	7	7	8	
	8	8	8	8	8	9	
	9	9	9	9	9	10	
	10	10	10	10	10	11	
		11	11	11	11	12	
		n.a	n.a	n.a	n.a	n.a	
	11	12	12	12	12	13	

2. Mapping Variable dan Jadual Pembuatan 2000-2013_Rev1.xlsx - Microsoft Excel

Data set variable-8.7.09-03.xlsx [Compatibility Mode] - Microsoft Excel

RECORD LAYOUT CONTENTS

NO.	CONTENT	2007	2005-2006	2004	2001-2003	2000 rebase	1983-2000(b4 rebase)	2007 File Size - 2007	2007 position	2005-2006 File Size - 2005-2006	2005-2006 position	2004 File Size - 2004	2004 position	2001-2003 File Size - 2001-2003	2001-2003 position	2000 (rebase) File Size - 2000 (rebase)	2000 (rebase) position	3-2000 (before reba File Size - 1983-2000(b4 rebase)	3-2000 (before reba position
28	Taraf Pendidikan	✓	✓	✓	✓	✓	✓	2	54	2	51	2	51	2	51-52	2	49-50	2	49-50
29	Sijil Tertinggi	✓	✓	✓	✓	✓	✓	2	56	2	53	2	53	2	53-54	2	51-52	2	51-52
30	Bidang Pengajian	✓	✓	✓	[redacted]			2	58	2	55	2	55	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
31	Institut Pengajian	✓	[redacted]					2	60	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
32	No. Ahli 15 tahun ke atas	✓	✓	✓				2	62	2	58	2	58	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
33	Tempat Kerja	✓	✓	✓	✓	✓	✓	2	64	2	60	2	60	0	60-61	2	58-59	2	58-59
34	Q1- Bekerja sekurang-kurannya sejam dalam minggu rujukan	✓	✓	✓	✓	✓	✓	1	66	1	62	1	62	1	62	1	60	#N/A	#N/A
35	Q2- Ada pekerjaan untuk dikembalikan ?	✓	✓	✓	✓	✓	✓	1	67	1	63	1	63	1	63	1	61	#N/A	#N/A
36	Q3- Bilangan jam bekerja?	✓	✓	✓	✓	✓	✓	3	68	3	64	3	64	3	64-66	3	62-64	#N/A	#N/A
37	Q4-Mengapa bekerja kurang dari 30 jam?	✓	✓	✓	✓	✓	✓	1	71	1	67	1	67	1	67	1	65	#N/A	#N/A
38	Q5- sanggup dan boleh menerima tambahan jam bekerja?	✓	✓	✓	✓	✓	✓	1	72	1	68	1	68	1	68	1	66	#N/A	#N/A
39	Q6- mengapa tidak bekerja dalam minggu rujukan?	✓	✓	✓	✓	✓	✓	1	73	1	69	1	69	1	69	1	67	#N/A	#N/A
40	Q7- mencari kerja minggu lalu?	✓	✓	✓	✓	✓	✓	1	74	1	70	1	70	1	70	1	68	#N/A	#N/A
41	Q8- kenapa tidak mencari pekerjaan pada minggu lalu?	✓	✓	✓	✓	✓	✓	2	75	2	71	2	71	2	71-72	2	69-70	#N/A	#N/A
42	Q9-Apakah anda buat minggu lalu?	✓	✓	✓	✓	✓	✓	1	77	1	73	1	73	1	73	1	71	#N/A	#N/A
43	Q10- Apakah langkah-langkah yang telah	✓	✓	✓	✓	✓	✓	5	78	5	74	5	74	5	74-78	5	72-76	5	72-76

ISSUE & CHALLENGES

DATA MANAGEMENT

System / Platform

Different versioning found in:

- **Data format used.**
- **Different application.**
- **Different system.**

Need data conversion before data can be migrated into DW.

ISSUE & CHALLENGES

DATA MANAGEMENT

Generation Issue

Generation issues found in :
- **Questionnaire ;**
- **Base year ;**
- **Code and Classification.**

Area

Country, State, District, *Mukim*, EB

Social

- **Ethnic**
- **Religion**
- **Nationality**
- **Type of houses**
- **Relationship to the head of HH**
- **Occupation status**
- **Field of study**
- **Classification of Disease & health care**
- **etc.**

Economy

- Malaysia Standard Classification of Occupations (MASCO)
- Malaysia Standard Industrial Classifications (MSIC).
- Malaysia Classification of Products by Activity (MCPA)
- Classifications of Functions of Government (COFOG).
- Standard International Trade Classification (SITC).
- Harmonized Commodity Description & Coding System (HS) .
- ASEAN Harmonized Tariff Nomenclature (AHTN).
- Classification of Individual Consumption by Purpose (COICOP).
- National Account Industrial Classification
- etc.

DATA MANAGEMENT

Where & Which the Final Data?

- Which data final at what level?
- Who keep the final data?

To have Standard Operation Procedure :
Where should we keep the real final data.

StatsDW OUTCOME

Improve Service Delivery System

- Accessibility
- Efficiency
- Increasing the capacity of data mining

Data Security

To protect the interest of all parties that rely on information systems from the effects of failure or weakness in terms of confidentiality, integrity, availability, accuracy of information and communication

Effective Data Management

- Type of census/ survey
- Frequency of census/ survey
- Generation of code and classification
- Generation of Structure and layout
- System and platform

WAY FORWARD

- Standard Operating Procedure
- User Acceptance
- Complete and comprehensive Data Management method and workflow

Thank You

Statistics Data Warehouse

For Better Data. Better Decisions.