

Development of ICT in Garnering Statistics

Sabri Omar

Department of Statistics Malaysia

Abstract

Statistical organisations all over the world are facing increased demand to raise quality, often through innovative approaches towards modernising the office environment and operations. Technological advancement in ICT has provided the required platform for Department of Statistics Malaysia (DOSM) to move forward. The Generic Statistical Business Process Model (GSBPM) adopted by DOSM has always been the reference in modernising our activities through ICT. The development of National Enterprise-Wide Statistical System (NEWSS) has created the better governance and housekeeping of enterprise and establishment frame as well as the household and housing frame. Perhaps NEWSS has enabled DOSM to monitor the progress of field work activities and the entry of collected information into the system. Besides that, intelligent character recognition, online data entry and e-Survey initiatives have improved the method of data capture. Current efforts on leveraging computer assisted telephone interviewing in conducting household surveys and computer assisted personal interviewing using computer tablets during field work are showing some benefits and are on the expansion.

The development of StatsDW data warehousing, StatsBDA big data analytics, open data, mobile applications are all geared towards new innovative initiatives leveraging on ICT technology.

1. Introduction

This paper describes the development of ICT initiatives in Department of Statistics Malaysia (DOSM) in garnering statistics.

DOSM has embarked into the use of ICT in its business function as early in 1962 [3][4]. The inception of ICL mainframe computer system has deliberately enhanced the processing of data collected during field work. Since then, ICT became an integral part of the core functions of DOSM.

2. General Statistical Business Process Model

The General Statistical Business Process Model (GSBPM) [5] adopted by DOSM has always been the reference in modernising the core functions through ICT. The 9-step GSBPM is shown in Table-1.

Table-1 : Generic Statistical Business Process Model (GSBPM)

Steps	Activities	Sub-activities
1	Specify Needs	1.1 Determine needs for information 1.2 Consult and confirm needs 1.3 Establish output objectives 1.4 Identify concepts 1.5 Check data availability 1.6 Prepare business core
2	Design	2.1 Design outputs 2.2 Design variable descriptions 2.3 Design data collection methodology 2.4 Design frame and sample methodology 2.5 Design statistical processing methodology 2.6 Design production systems and workflow
3	Build	3.1 Build data collection instrument 3.2 Build or enhance process components 3.3 Configure workflows 3.4 Test production system 3.5 Test statistical business process 3.6 Finalise production system
4	Collect	4.1 Select sample 4.2 Set up collection 4.3 Run collection 4.4 Finalise collection
5	Process	5.1 Integrate data 5.2 Classify and code 5.3 Review, validate and edit 5.4 Impute 5.5 Derive news variables and statistical units 5.6 Calculate weight 5.7 Calculate aggregate 5.8 Finalise data file

Steps	Activities	Sub-activities
6	Analyse	6.1 Prepare draft output 6.2 Validate outputs 6.3 Scrutinise and explain 6.4 Apply disclosure control 6.5 Finalise output
7	Dissemination	7.1 Update output system 7.2 Produce dissemination products 7.3 Manage release of dissemination products 7.4 Promote dissemination products 7.5 Manage user support
8	Archive	8.1 Define archive rule 8.2 Manage archive repository 8.3 Preserve data and associated metadata 8.4 Dispose of data and associated metadata
9	Evaluation	9.1 Gather evaluation input 9.2 Conduct evaluation 9.3 Agree action plan

Mapping of ICT initiatives against the GSBPM [11] is depicted in Table-2.

Table-2 : GSBPM and ICT Initiatives

Steps	Activities	ICT Initiatives
1	Specify Needs	<ul style="list-style-type: none"> • Feasibility studies and Edit Specifications
2	Design	<ul style="list-style-type: none"> • National Enterprises-Wide Statistical System (NEWSS) • Statistical packages
3	Build	<ul style="list-style-type: none"> • NEWSS
4	Collect	<ul style="list-style-type: none"> • Intelligent Character Recognition (ICR) • Computer Assisted Personal Interviewing (CAPI) - using Tablet • Computer Assisted Telephone Interviewing (CATI) • e-Survey • Online Data Entry
5	Process	<ul style="list-style-type: none"> • NEWSS • StatsDW Data Warehouse • StatsBDA Big Data Analytics
6	Analyse	<ul style="list-style-type: none"> • SAS • SPSS • eView • Stata • MS Excel
7	Dissemination	<ul style="list-style-type: none"> • Portal • Visualization • Geographical Information System (GIS) • Stats Data Warehouse (StatsDW) • Stats Big Data Analytics (StatsBDA)

Steps	Activities	ICT Initiatives
8	Archive	<ul style="list-style-type: none"> • StatsDW
9	Evaluation	<ul style="list-style-type: none"> • Manual

These ICT initiatives are further categorised as follows:

- a) Established ICT initiatives,
- b) Rolling out ICT initiatives, and
- c) Way forward ICT initiatives.

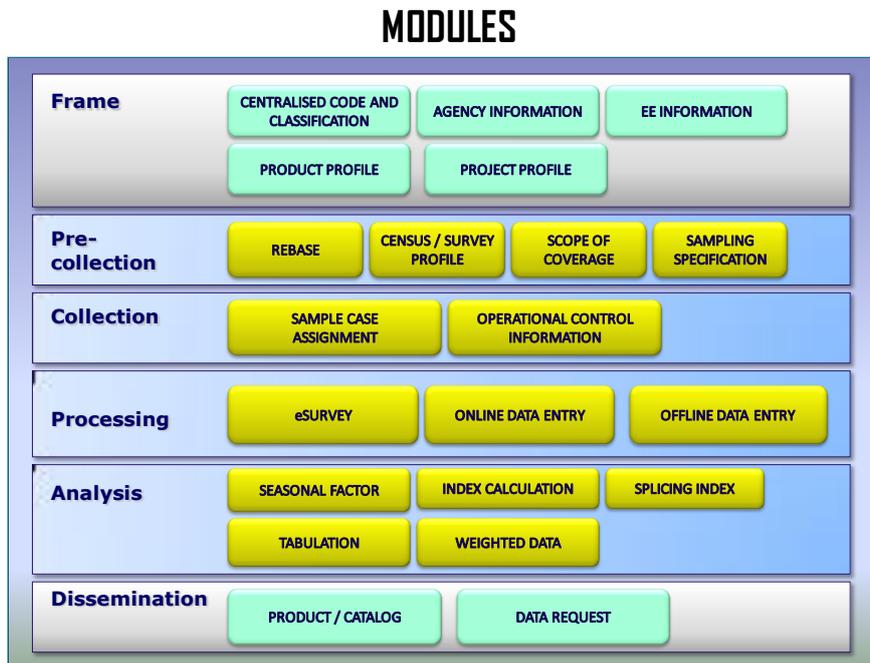
ESTABLISHED ICT INITIATIVES

3. National Enterprise-Wide Statistical System

The National Enterprise-Wide Statistical System (NEWSS) was developed using the Fujitsu Framework introduced by Fujitsu (M) Private Limited company in 2008 [8]. This JAVA Framework comprises of the Integrated Statistical System Framework and Information System Support.

The modules of NEWSS are shown in Figure-1.

Figure-1 : Modules of NEWSS

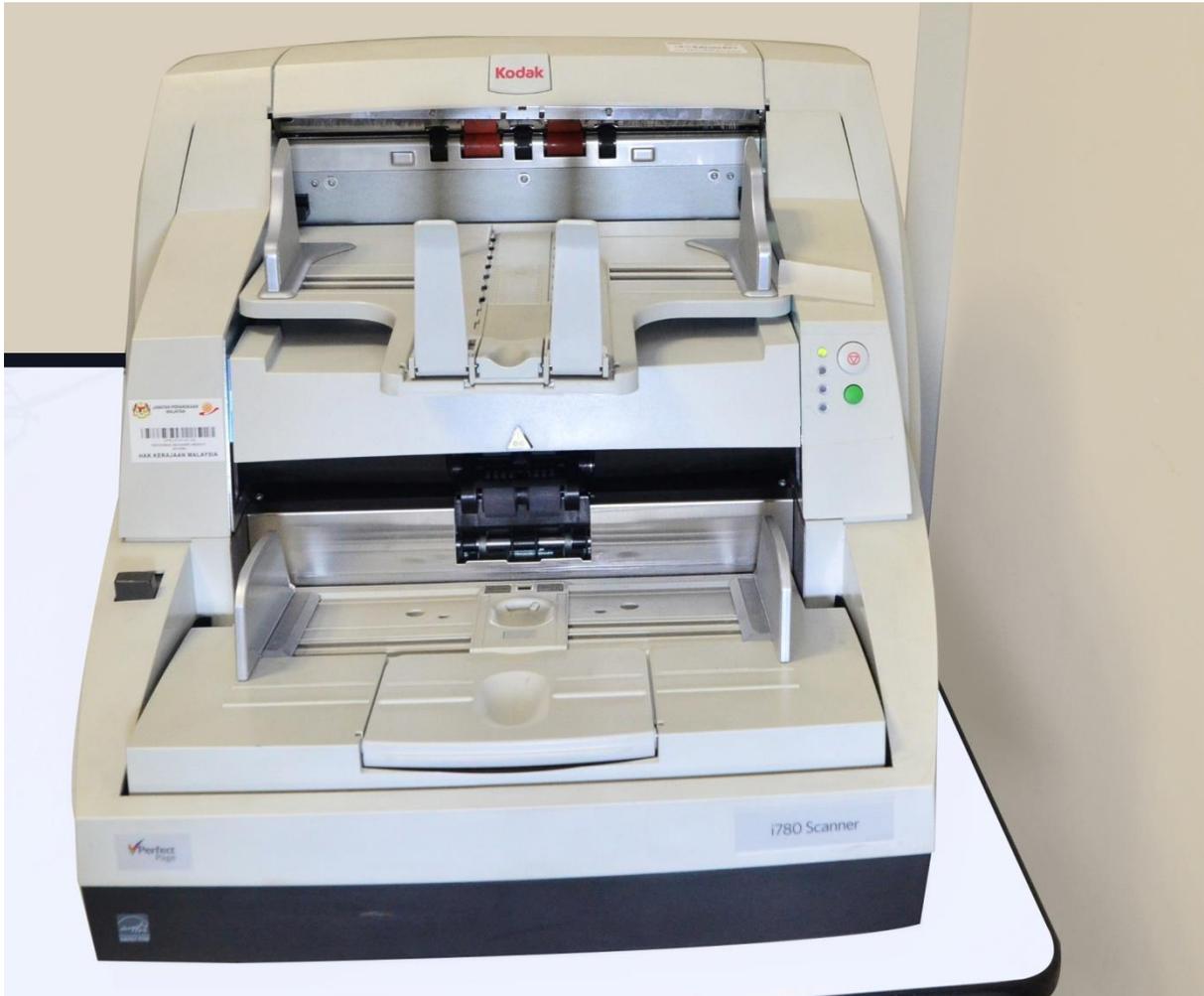


4. Intelligent Character Recognition

Intelligent Character Recognition (ICR) machine and software have been used intensively in 2005 Agriculture Census. It has helped in capturing data from the physical paper based questionnaires into Microsoft Sqlserver database through scanning, transfer and verify of the scanned images.

It tremendously reduced the time in data capture compared to manual key-in by clerical staffs. The current model of the scanner is Kodak i780 and the software is Readsoft version SP5-2.

Figure-2 : ICR Machine



Simple calculations to show the capability of the ICR Machine are as follows:

Recommended daily volume is 130,000 pages per day.

Throughput speed is 130 pages per minute.

Based on throughput speed, number of pages scanned per hour is $130\text{ppm} \times 60$ minutes equals 7,800 pages.

If the machine scanned for 8 hours per day, the number of pages scanned per day is $7,800$ pages \times 8 hours equals 62,400 pages per day.

Assuming a set of questionnaires has 12 printed pages the number of sets of questionnaires scanned is $62,400 / 12$ equals 5,200 (sets of questionnaires).

If there are 80,000 sets of questionnaires, the number of days required to scan is $80,000 / 5,200$ equal 15.4 days.

Based on the calculations, the ICR machine is able to speed up data entry.

5. Online eSurvey

Modernising the date of collection method has been an important agenda in the ICT initiative program. Online eSurvey has been implemented since 2011 and currently there are six online eSurvey and their performing status are shown in Table-3.

Table-3 : Online eSurvey

System	No. of Respondent	eSurvey	Other Electronics Form (Email, Fax & Telephone)	Total eSurvey of other electronics form
Monthly Manufacturing (eMM) (2011)	3,321	721 (21.7%)	1,820 (54.8%)	2,541 (76.5%)
Quarterly Services Survey (eQSS) (2012)	8,352	387 (4.6%)	3,291 (39.4%)	3,678 (44%)
International Trade in Services (eITS) (2013)	489	235 (48.1%)	205 (41.9%)	440 (89.9%)
eJobCreation (2014)	5,598	443 (8%)	3,286 (58.7%)	3,729 (66.6%)
Monthly Distributive Trade (eMDT) (2014)	4,372	18 (0.4%)	3,140 (71.8%)	3,158 (72.2%)
Quarterly Construction Survey (eQCS) (2015)	4,718	18 (0.4%)	2,953 (62.6%)	2,971 (62.9%)

Source : DOSM Main User Committee, 2015

The combination of responses via eSurvey and other electronic form showed the potential acceptance of respondents using eSurvey. However substantial effort are needed to encourage the respondents to fully use eSurvey.

6. e-Services Portal

e-Services is an online system provided for the convenience of users to get the latest DOSM products and services. e-Services offers facilities such as user registration, free download publications, Advance Release Calendar, subscription of notification email for latest publication, purchase or subscribe publications, data requests for unpublished data, review transactions status, online payment systems via credit card, feedback and selected DOSM e-survey [6].

The benefits of e-Services are as follows:

- a) Seamless accessibility available 24 hours, 7 days a week.
- b) Personalised services that facilitates users.
- c) Hassle-free services
- d) Free Downloads Statistical Products
- e) Cashless Transactions

Figure 3 shows the e-Services login page.

Figure-3: e-Services Login Page



Figure 4 shows the list of free download publication whilst and Figure 5 shows the online payment via credit card

Figure-4: Free download publications

DEPARTMENT OF STATISTICS MALAYSIA
e-Services

Home

My Account

Publications

- Printed
- Free Download
- Journals

Advance Release Calendar

- Advance Release Calendar

e-Survey

- MM e-Survey (Monthly Manufacturing Survey)
- MDT e-Survey (Monthly Distributive Trade Survey)
- QCS e-Survey (Quarterly Construction Survey)
- PPI e-Survey (Monthly Price Survey)

Data Request & Feedback

- Order Form
- Feedback Form

User Guide

- e-Services
- QCS e-Survey
- MM e-Survey

Search | Advance Search

Free Download

Main Category: Please select

Sub-Category: Please select

Search Reset

No.	Title of Publications	Product Type	Release Series	Release Date
1	Statistics of Foreign Direct Investment In Malaysia	Publication	Year 2014	30 October 2015
2	Abridged Life Tables, Malaysia	Publication	Year 2015	30 October 2015
3	Producer Price Index Malaysia - Local Production	Publication	September 2015	30 October 2015
4	Import & Export Price Indices, Malaysia	Publication	September 2015	30 October 2015
5	Monthly Statistical Bulletin, Malaysia	Publication	August 2015	30 October 2015
6	National Accounts Gross Domestic Product (GDP) By State	Publication	Year 2015	30 October 2015
7	Consumer Price Index, Malaysia	Publication	September 2015	23 October 2015
8	Malaysia Economic Indicator - Leading, Coincident and Lagging Indices	Publication	August 2015	23 October 2015
9	Report on the Survey of Environmental Protection Expenditure, Malaysia	Publication	Year 2014	20 October 2015
10	Monthly Rubber Statistics, Malaysia	Publication	August 2015	12 October 2015

Page: 1 2 3 4 5 6 7 8 9 10 » »

Record 1 to 10 from 1,376

Figure-5: Online Payment via Credit Card

BANK ISLAM

Merchant name: JABATAN PERANGKAAAN MALAYSIA

Select your preferred payment method

Pay securely using SSL+ by clicking on the card logo below:

VISA MasterCard

Copyright ©2007 TNS Payment Technologies Pty Ltd. All Rights Reserved.

SECURE PAYMENTS POWERED BY DIALECT

7. Geographical Information System

A geographical information system (GIS) is a system designed to capture, store, manipulate, analyze, manage, and present all types of spatial or geographical data. In DOSM, NEWSS GIS is used for delineation and digitizing of

Enumeration Block (EB) boundaries. NEWSS GIS consist of two sub- modules, which is check in check out (CICO) that involves updating EB by State Offices and approval process by Headquarters and GIS Portal as a platform to display data in the form of thematic maps. It also provides a search facility for household and establishment based on the selection area which is generated directly from NEWSS Household and Establishment frame.

Besides that, DOSM also published Census e-Atlas 2010 for public [6]. Census e-Atlas 2010 was developed for the first time using data obtained from the Population and Housing Census of Malaysia 2010. It is a mechanism that aims to show the main theme of the Census in the form of thematic maps at the national and state levels. The themes covered are the distribution of population, ethnicity, religion, age structure and marital status.

Figure-6: NEWSS GIS

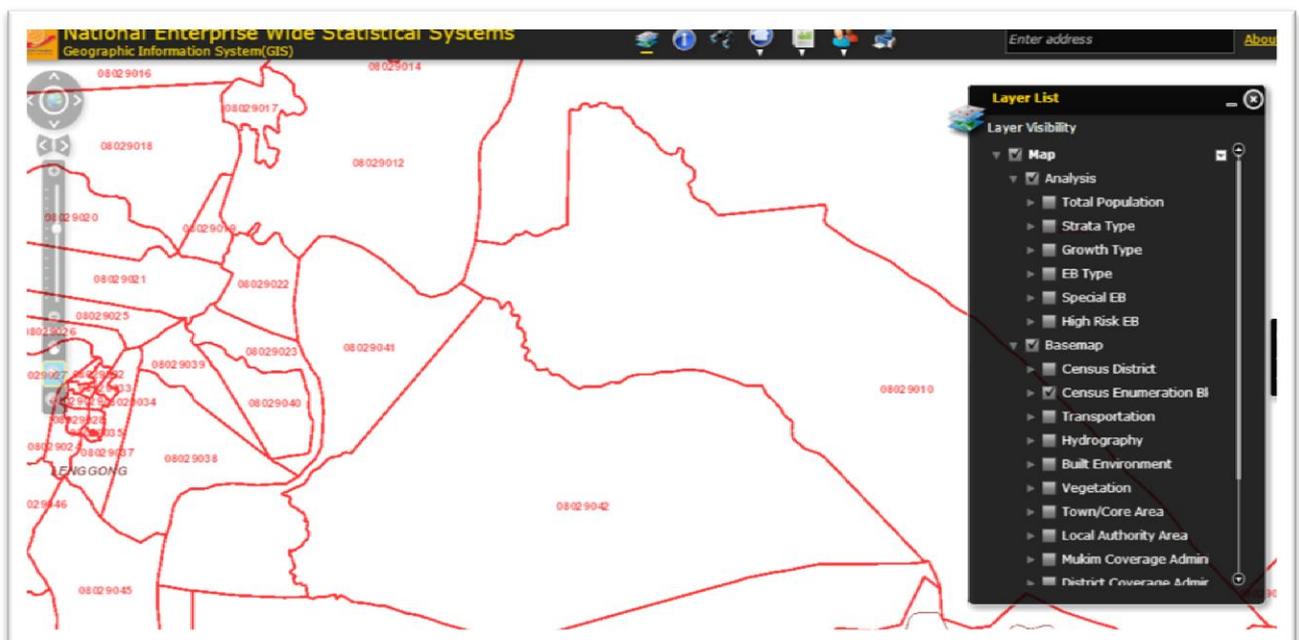
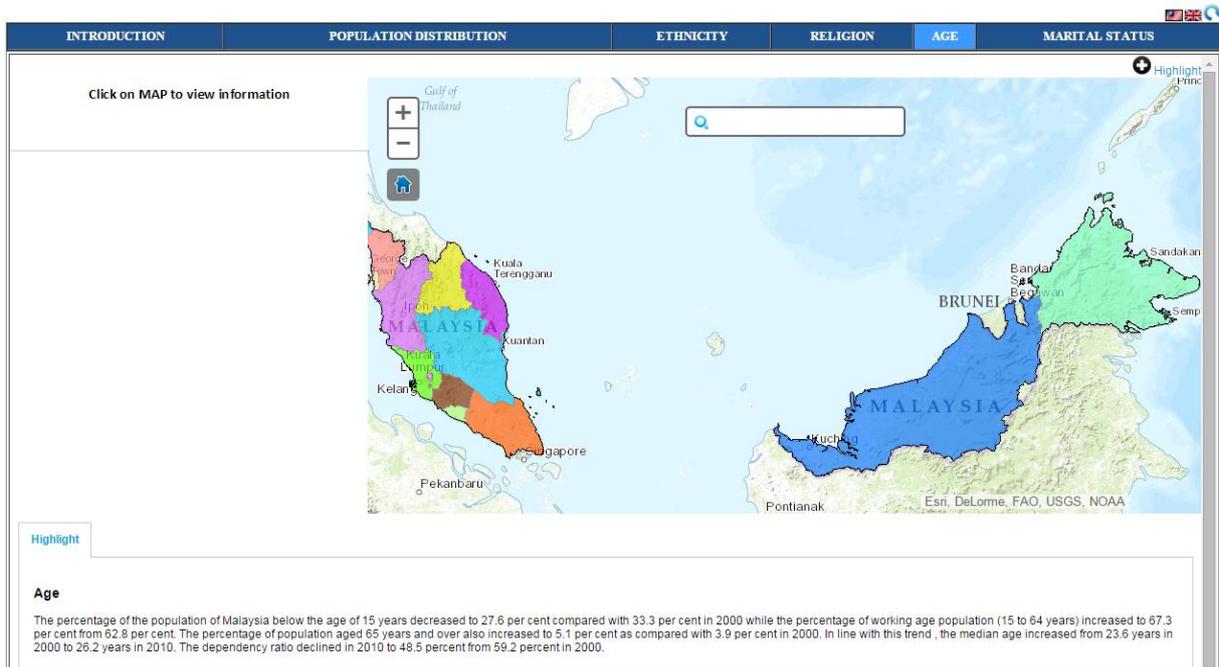


Figure-7: Census e-Atlas 2010



ROLLING OUT ICT INITIATIVES

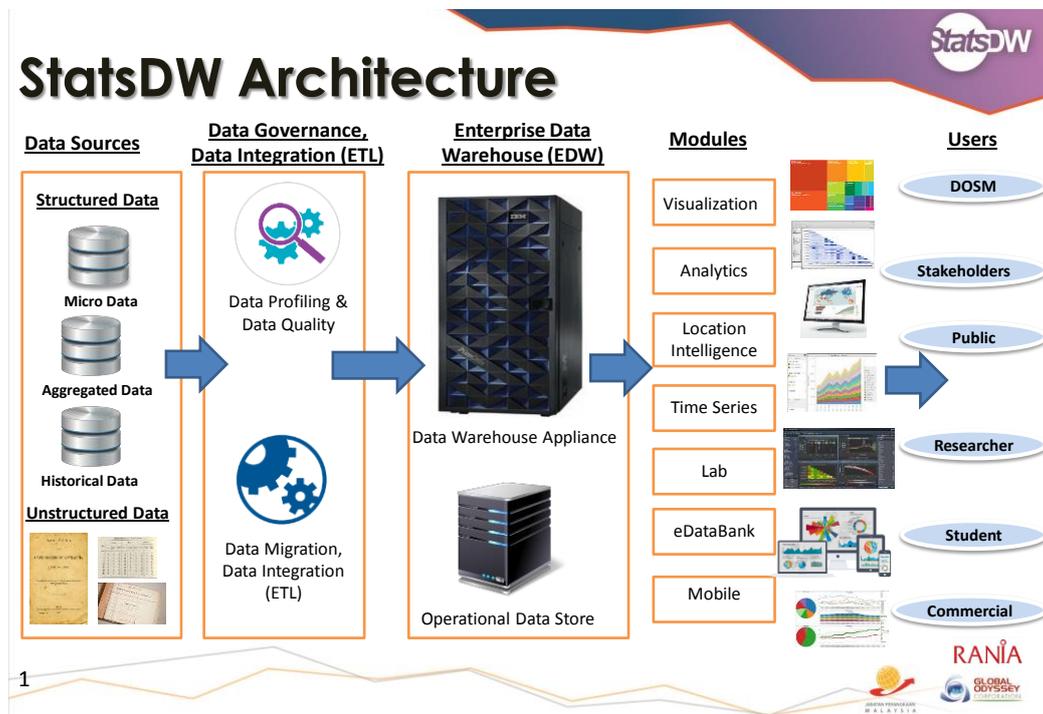
8. StatsDW Data Warehouse

StatsDW data Warehouse is a relational database that is designed for query and analysis rather than for transaction system. It usually contains historical data derived from transaction data but can include data from other sources. It separates analysis workload from transaction workload and enable organisation to consolidate data from several sources.

In addition to a relational database, a data warehouse environment includes an extraction, transportation, transformation, and loading (ETL) solution, an online analytical processing (OLAP) engine, client analysis tools, and other application that manage the process of gathering data and delivering it to business users [1].

Figure-3 shows the StatsDW Data Warehouse model.

Figure-8 : StatsDW Data Warehouse Model



The modules that are developed under StatsDW are as follows [13,14]:

- a) Visualization
- b) Analytics
- c) Location Intelligence
- d) Time Series
- e) StatsDW Laboratory
- f) e-Data Bank
- g) Mobile Applications.

9. Computer Assisted Telephone Interviewing

Another modernisation initiative is the use of computer assisted telephone interviewing (CATI) technique in which the interviewer follows a script provided by a software application. It is a structured system of microdata collection by

telephone that speeds up the collection and editing of microdata and also permits the interviewer to educate the respondents on the importance of timely and accurate data [2].

Currently, DOSM has set up three CATI centres which are located at the headquarters, Putrajaya; Federal Territory Kuala Lumpur Department of Statistics which serves as the centre for central zone covering respondents in Kuala Lumpur and Selangor; and Department of Statistics Melaka which serves the southern zone covering Negeri Sembilan, Melaka and Johor.

Based on the comparative study by Department of Statistics Federal Territory Kuala Lumpur, there is cost saving in the use of CATI compared to the normal face to face interview with the respondents. Table-4 shows the cost saving of CATI for Labour Force Survey.

Table-4 : Cost Saving of CATI for Labour Force Survey

Assumptions:

No. of cases	:	4,416 cases @ 552 EB
No. of Staff	:	39 staffs (Frame, Field work, Processing)
KPI for Field work	:	2 cases per day per staff
KPI for CATI	:	7 cases per day per staff

	Face to Face Interviewer (RM per year)	CATI (RM per year)
Frame	103,541	103,541
Field Work	391,521	197,017
Processing	288,000	216,000
Total	783,062	516,558

Source: DOSM Federal Territory, Kuala Lumpur

Based on the above findings, the cost saving are as follows:

- a) Cost saving for field work is 49.7%
- b) Cost saving for processing is 25%
- c) Total cost saving is RM266,504.

The benefits gained in the use of CATI have encouraged to the set-up of new centres for northern zone in Penang, eastern zone in Pahang, Sabah zone and Sarawak zone.

10. Computer-Assisted Personal Interviewing

Computer-assisted personal interviewing (CAPI) is an interviewing technique in which the respondent or interviewer uses a computer to answer the question. It is similar to computer-assisted telephone interviewing, except that the interview takes place in person instead of over the telephone. It has been classified as a personal interviewing technique because an interviewer is usually present to serve as a host and to guide the respondent [16].

For CAPI, DOSM has introduced the use of tablet for Customer Price Index (CPI) data collection in five states, that is Federal Territory Kuala Lumpur, Selangor, Johor, Perak and Penang. The tablet is used to replace the manual book in which Field Enumerator has to fill in during field work and re-key-in into CPI system for validation when return to office.

On the other hand, data are key-in directing into the tablet during field work and uploaded into the CPI system when return to office. The rekey-in of the data is not necessary anymore. Table-5 shows the benefits of using of CAPI.

Table-5 : Benefits of using CAPI for CPI

Process	Book Manual	CAPI
<p><u>In the Office</u></p> <p>1. Preparation of Book / File</p> <p>2. Duration of preparation</p> <p>3. Printing cost per year (paper, toner, overtime)</p> <p><u>Field Work</u></p> <p>1. Data Collection</p> <p>2. Action code</p> <p><u>Data Processing</u></p> <p>1. No. of Staff</p> <p>2. Time to process</p> <p>3. Data check and Quality</p>	<ul style="list-style-type: none"> • Long time to print • Copy previous price manually • 5 days • RM21,390 • Book is big and heavy • Action code error cannot be detected • 16 staff • 2~3 days • Long time because need to refer CPI Book 	<ul style="list-style-type: none"> • Check 3 Excel Files-Shop, Item non-spec, month • Consolidated in the system • 2~3 days • RM1,937 (saving of RM19,453) • Light and user friendly • System alert if data item is incomplete • System detects action code error • 2~3 staff • 1 day • Fast because check by system

Source: DOSM Selangor

11. Mobile Application

Widely use of smart phones and tablets have triggered the development of mobile applications or commonly known as mobile apps. Price check mobile application provides the public to search and compare monthly average price of products for certain category and location.

It provides the average price of product for the state; highest and lowest price for the current month; and comparing the average price between the current and previous month.

More mobile apps are to be developed and that includes the search engine for public to search statistical transactions.

WAY FORWARD ICT INITIATIVES

12. Open Data

By definition, open data means that anyone can access, use or share. Simple as that. When big companies or governments release non-personal data, it enables small businesses, citizens and medical researchers to develop resources which make crucial improvements to their communications [10].

For Malaysia, the open data initiative through the portal data.gov.my is spearheaded by Multimedia Development Corporation (MDeC) and Malaysia Administrative Modernisation and Management Planning Unit (MAMPU).

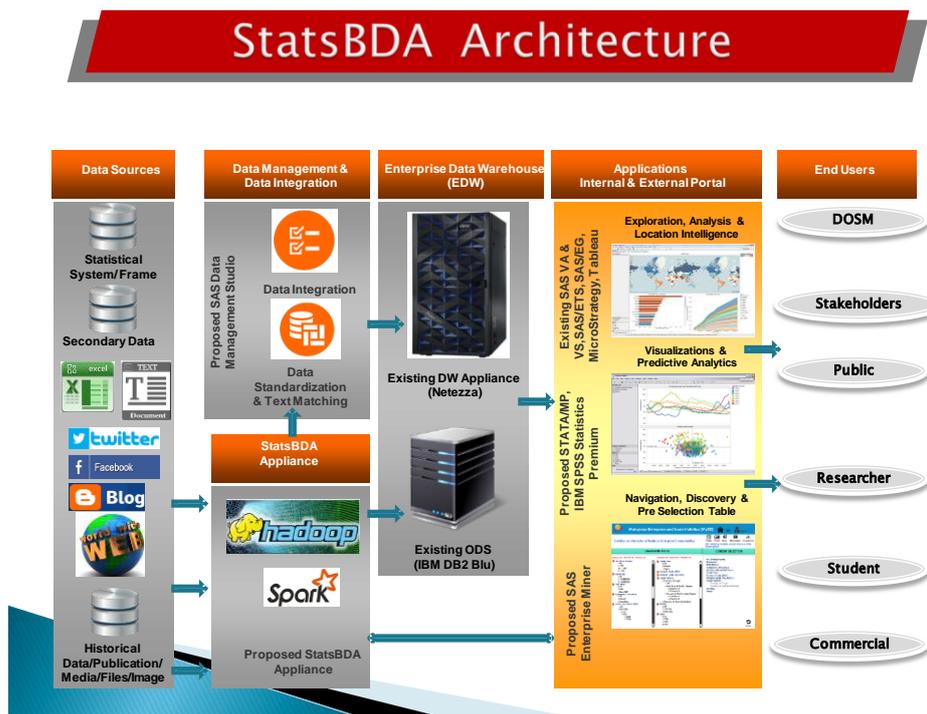
This open data initiative is still new and serious thought and planning are undergoing with many government agencies, including DOSM. The ultimate open data must be in comma-separated values (CSV) file that stores tabular data (number and text) in plain text. Each line of the file is a data record. Each record consists of one or more fields, separated by commas. The use of the comma as a field separator is the source of the name for this file format [15].

13. Big Data Analytics

Big data is being generated by everything around us all the times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Big data is arriving from multiple sources at an alarming velocity, volume and variety. To extract meaningful value from big data, you need optimal processing power, analytics capabilities and skills [7].

DOSM has embarked on a journey of big data by devising its big data model, named StatsBDA model. Figure-9 depicts the StatsBDA model.

Figure-9 : StatsBDA Model



The phase 1 of the StatsBDA implementation will involve the time series data of Malaysian External Trade, Malaysia statistical business register, and Malaysia Statistical address register.

14. Development of Application Framework

Current system development process uses the conventional system development life cycle (SDLC) which has seven stages namely project planning and feasibility study; system analysis and requirements definition; system design; implementation; integration and testing; acceptance, installation and deployment; and maintenance [12].

However, with the high demand and expectation from the users, stages of the SDLC need to be innovated by adopting system development framework. Furthermore, the development of web 2.0 application requires Yii as the high performance PHP Framework [9].

15. Conclusion

In this paper, the development of ICT initiatives have covered the three categories of established, rolling out and way forward ICT initiatives. These initiatives have significantly helped DOSM to modernise its core functions to produce statistical data and publications.

The ICT initiatives have shown benefits in terms of monetary value and time. However, some initiatives need further study on the benefit realization in order to measure the outcomes.

16. Acknowledgements

I would like to express my gratitude to DOSM Federal Territory Kuala Lumpur - Datin Rozita binti Talha and Cik Sa'adiyah binti Ahmad for input on CATI; DOSM

Selangor - Tuan Haji Ibrahim bin Jantan for input on CAPI; and Mr. Mior Norazman bin Mior Daud, for input on GIS and e-Services. My gratitude also to Cik Nazlah binti Mustaffa for the infographics and Cik Nazirah binti Nazarudin for the type setting of this paper.

Tables and Diagrams

Table

1	Table 1	:	Generic Statistical Business Process Model
2	Table 2	:	GSBPM and ICT Initiatives
3	Table 3	:	Online e-Survey
4	Table 4	:	Cost Saving of CATI for Labour Force Survey
5	Table 5	:	Benefits of using CAPI for CPI

Diagram

1	Figure 1	:	Modules of NEWSS
2	Figure 2	:	ICR Machine
3	Figure 3	:	e-Services Login Page
4	Figure 4	:	Free download publications
5	Figure 5	:	Online Payment via Credit Card
6	Figure 6	:	NEWSS GIS
7	Figure 7	:	Census e-Atlas 2010
8	Figure 8	:	StatsDW Data Warehouse Model
9	Figure 9	:	StatsBDA Model

Bibliography

1. Building the Data Warehouse, William Inmon (John Wiley and sons), 1996
2. CATI : Glossary U.S. Bureau of Labour Statistics
Division of Information Services
3. DOSM 60 Year Book, 2009
4. DOSM Annual Report, 2014
5. DOSM - General Statistical Business Process Model, 2011
6. DOSM - Portal <https://www.statistics.gov.my>, 2015
7. IBM : What is Big Data?, 2015
8. Fujitsu Framework - Development Guidelines & Standard, 2008
9. Mark Safronov, Jeffery Winesett, Web Application Development with Yii 2 and PHP, 2014
10. Open Data Institute, United Kingdom, 2015
11. Presentation slides in Statistical Process Modernisation Committee, 2015
12. Rusell Kay, System Development Life Cycle, Computerworld, 2002
13. StatsDW DOSM - Functional Design Specifications, 2014
14. StatsDW DOSM - User Requirement Specifications, 2014
15. Wikipedia : Comma-separated values, 2015
16. Wikipedia : Computer-assisted personal interviewing, 2015