Dashboard on Social & Economic Statistics: A Case Study in Mukim Ayer Hitam, Muar, Johor, Malaysia

Riyanti Saari and Suraimey Sulaiman[1]

Abstract

Data integration is very important and powerful in bringing together data from different sources and gaining more valuable outcome. Thus, the Department of Statistics Malaysia (DOSM) has developed a dashboard as a platform to integrate social and economic statistics with spatial data.

The first case study of self-updating on basic information of living quarters (LQ), population and integration with Malaysia Statistical Business Register (MSBR) was held in Mukim Ayer Hitam, Johor, Malaysia in June 2022.

It comprised 22 Enumeration Blocks, 2,364 geocode of each LQs, 6,798 people and 150 establishments with geocodes. This Dashboard is also integrated with other basic facilities, such as schools, hospitals / clinics, homestays and natural disaster areas (potential flood areas).

This case study is undertaken by the DOSM in collaboration with district office of Muar, Johor, Malaysia. It has been extended to other districts or localities in all states. It also serves as one of the inputs for the updating the Malaysia Statistical Address Register (MSAR) and Malaysia Statistical Population Profile (MSPP).

Keyword: data integration, spatial data, social statistics and economic statistics

[1] Riyanti Saari is currently Director of Methodology and Research Division and Suraimey Sulaiman is currently the Senior Assistant Director of Methodology and Research Division, Department of Statistics Malaysia.

Introduction

Statistical business register (SBR) is a registry of enterprises or establishments involved in economic activities. Malaysia has also developed its own SBR known as Malaysia Statistical Business Register (MSBR). It is a fundamental property in maintaining the comprehensive list of businesses and companies operating in Malaysia which encompasses Registration of Companies (ROC), Registration of Business (ROB) and Limited Liability Partnership (LLP) that are registered with Companies Commission Malaysia (CCM) as well as establishments that are registered with local authorities and professional bodies. The main source of this registry is mainly from the Companies Commission of Malaysia. In addition, DOSM also collaborates and collects information from other agencies, such as the Employee Provident Fund, the Royal Malaysian Customs Department, Inland Revenue Board of Malaysia, local authorities and other professional bodies that is used to update MSBR. Surveys and censuses conducted by DOSM is also used as a source to update enterprises and establishments information in MSBR.

MSBR is very important in the establishment of a common framework in the creation of harmonised registers for statistical purposes. It provides a survey frame and profile of establishments with basic characteristics, such as economic activity in terms of revenue, employment size, establishment size (large, medium, small and micro) statistics and local kind of activity units as well as enterprise groups, central and local government bodies. It is a basic pillar which facilitate the collection, interpretation and dissemination of economic statistics. Establishments are classified to their industries based on Malaysia Standard Industrial Classification (MSIC) 2008 Version 1.0 which is comparable to ISIC 4 of the United Nations. MSBR can be more useful if it is integrated or linked to geospatial information. Thus, in order to enrich the existing database of MSBR, the linkages of geospatial information to each establishment in MSBR is developed.

This paper aims to present the pilot study on integration of social and economic statistics with spatial data in Mukim Ayer Hitam, Muar, Johor, Malaysia. The result of the case study is visualised through a dashboard to generate the interest of the public and to showcase the valuable and comprehensive information made possible through data integration. This paper will also briefly describe the development of spatial information for 2023 Economic Census's analysis and for 2024 Agriculture Census's preparation.

Methodology

The existence of MSBR is very important as a structured database for survey design and operation especially for all economic surveys and censuses. It provides sampling frame for establishment surveys. The number of registered establishments in Malaysia was about 1.2 million. There are five main sectors namely Agriculture, Mining & Quarrying, Manufacturing, Construction and Services. Manufacturing and Services sector that can be further disaggregated into subsectors. Some of the basic information in the MSBR are CCM number, company name, postal address and locality code. As a new innovative procedure, all the establishments in MSBR is going to be linked with the spatial information - geocode (X,Y) based on the establishment address. These linkages are very important and vital to ensure that small area statistics published are accurate. Spatial data can be referred to as geographic data or geospatial data and provides information that can be used to identify the location, features and boundaries of the Earth. It also can be processed and analysed using Geographical Information Systems (GIS) or Image Processing packages.

In line with the Global Statistical Geospatial Framework (GSGF), Principle 2, support the process of linking or storing high precision geographic references to each microdata/ statistical unit record. While, the goal of Principle 2 is to allow all statistical unit records to be linked to a location. This link will then enable integration of data from a wide variety of sources, such as other socio-economic statistical data, administrative data, and geospatial information for instance, the built, and natural environment.

Thus, strengthening the MSBR with spatial information is one of the priorities in order to produce reliable and accurate small area business statistics through data integration. Data integration refers to the process of bringing together data from

multiple sources across an organisation to provide a complete, accurate, and up-to-date dataset, data analysis and other applications and business processes. In the case study of Mukim Ayer Hitam, Muar, Johor, the integration process was done using data set of MSAR, MSPP, MSBR from DOSM as the first step before integration with information from other government agencies to DOSM as the first step before integration with latest information from self updating by Mukim Ayer Hitam community.

Geocoding Platform and Geocoding output

DOSM implemented two types of geocoding platforms which are Geocode Awesome @Google Sheet and MyGeoTranslator @PGN. Geocode Awesome @Google Sheet. It means that the longitude and latitude (X,Y) will be automatically generated. Meanwhile, for MyGeoTranslator @PGN, the matching locality is based on geocoded data (e.g. state, administrative district, mukim/locality and election boundaries (parliamentary and Subnasional Statistics of State Legislative Assembly).

Table 1: Geocode Awesome @Google Sheet

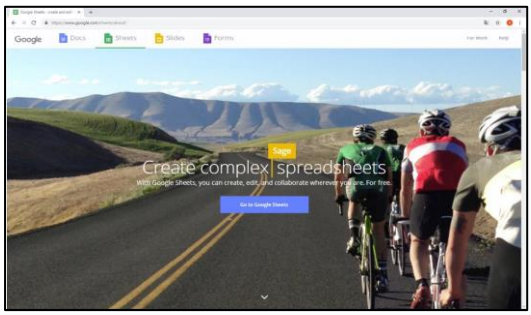
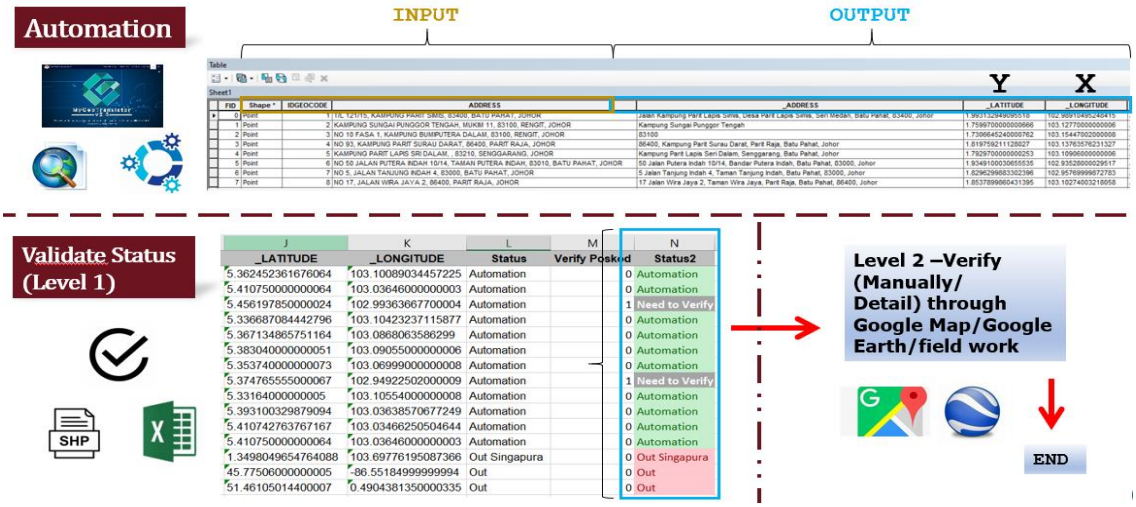


Table 2: MyGeoTranslator @PGN



Benchmark of Data Integration by Nordic Register Based Statistical System and Malaysia proposed

Figure 1: Nordic Register-based Statistical System

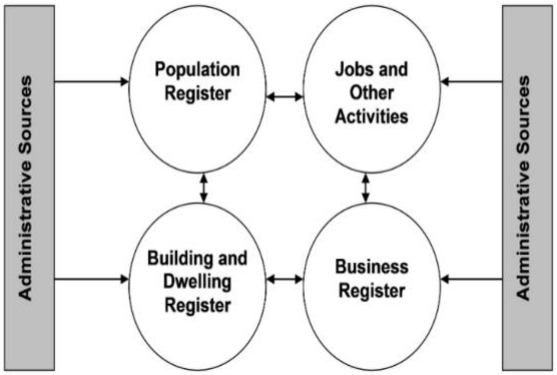
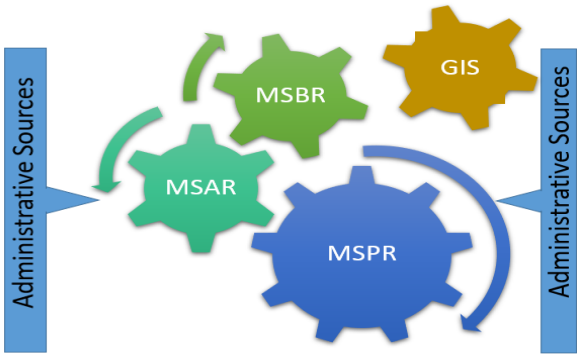


Figure 2: DOSM, MALAYSIA*: MSAR, MSBR, MSPR & GIS Integration*



In all Nordic countries, three administrative-based registries are defined, register on persons (population registry), registry on properties (real estate, buildings and dwellings) and business registry. While in Sweden, a fourth statistical based registry is defined as the activity registry (job, other labour market-related activities and educational activities).

In terms of address, Australia have developed the Address Registry contains addresses, address function (e.g. residential, commercial, etc.), a latitudinal and longitudinal coordinates known as a "geocode" and dwelling structure for residential addresses (e.g. separate house, semi-detached, etc.). Meanwhile, the Netherlands' population register system has been in use to cover every person residing in the Netherlands as a basic element in the governmental work processes at the national and local level.

For Malaysia, MSBR has been developed since 1980s, which is called as Central Registration System (CRS). Starting 2010, it was improved through National Enterprise Wide Statistical System and called MSBR. Meanwhile, through Malaysia Population & Housing Census (MYCENSUS) 2020, among the main outcomes are Malaysia Statistical Address Register (MSAR) and Malaysia Statistical Population Register (MSPR). Thus, the integration and linkages between MSBR, MSAR and

MSPR with spatial information, is very important in order to produce more accurate and quality statistics.

A case study of self-updating by community on basic information of living quarters (LQ) and the population was held in Mukim Ayer Hitam, Johor, Malaysia started in June 2022. It consists of 22 Enumeration Blocks, 2,705 geocodes of each LQs, 6,798 people and 150 geocodes of each establishment. This dashboard has also been integrated with basic facilities, such as schools, hospitals / clinics, homestays and natural disaster areas (potential flood areas). This case study is a collaboration between the Department of Statistics Malaysia and District Office of Muar, Johor, Malaysia.

The first step of integration is identifying the key indicator for each data sources. For MSBR, the main identifier is the CCM number. Meanwhile, for MSAR it is the Address ID and NGDBBP and for MSPP, Identification Card Number and Address ID are used as identifiers.

Figure 1: MSBR, MSAR and MSPP

MALAYSIA STATISTICAL BUSINESS REGISTER
ESTABLISHMENT_ID
BUSINESS_REG_NO
REGISTERED_NAME
TRADING_NAME
STREET_1_B
STREET_2_B
TOWN_B
POSTCODE_B
STATE_B
DISTRICT
MUKIM
SECTOR
MSIC 2008
YR_KATEGORI_EKS_PKS
NGDBBP
LONGITUDE
LATITUDE

MALAYSIA STATISTICAL ADDRESS REGISTER
ADDRESS_ID
NGDBBPTK
DISTRICT
MUKIM
BU_NO
LQ_NO
UNIT_NUMBER
LEVEL
BLOCK
BUILDING_NAME
STREET_NAME
SECTION_NAME
CITY_NAME
POSTCODE
STATE_CODE
LONGITUDE
LATITUDE

MALAYSIA STATISTICAL POPULATION PROFILE
ADDRESS_ID
NGDBBPTK
DISTRICT
MUKIM
BU_NO
LQ_NO
UNIT_NUMBER
LEVEL
BLOCK
BUILDING_NAME
STREET_NAME
SECTION_NAME
CITY_NAME
POSTCODE
STATE_CODE
IDENTIFICATION CARD NO.
NAME
RELATION
SEX
MARITAL STATUS
ETHNIC
RELIGION
EDUCATION LEVEL
OCCUPATION
SALARY
DISABLED PERSON
INVOLVED AGRICULTURE ACTIVITY

Discussion

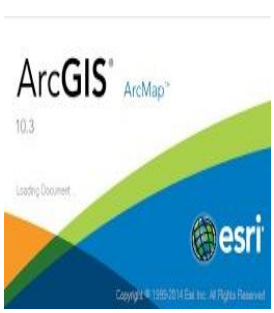
Step 1: Verification of geocode for each establishment and living quarters

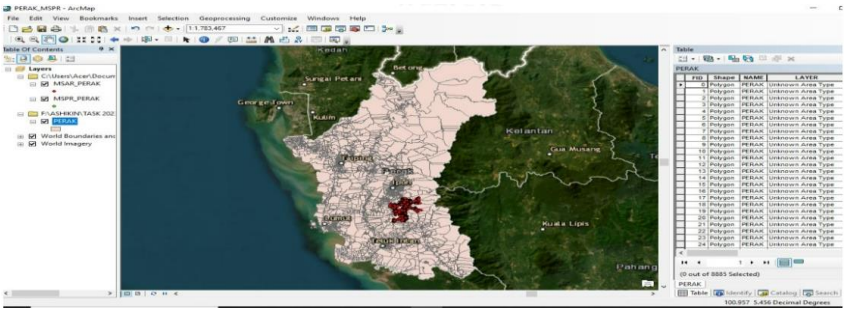
Geocode for each establishment and living quarters in Mukim Air Hitam will be verify by overlying with spatial layer (desktop canvassing). If there are any discrepancies, it will be checked and confirmed on the ground (field work) by DOSM State's Office.

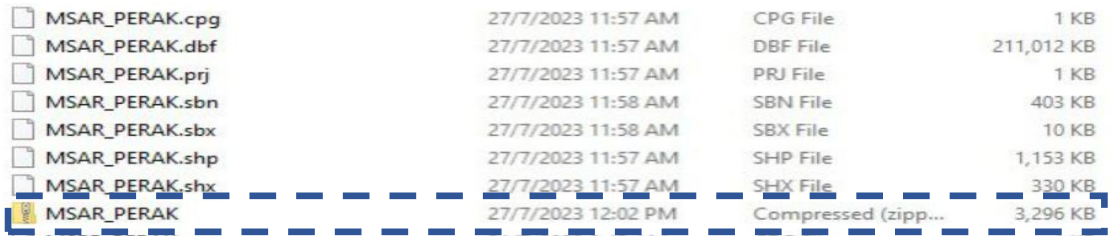
REGISTERED_NAME	STREET_1_B	STREET_2_B	TOWN_B	STATE_B	X_MYG	Y_MYG	NGDBBP202
HONG LEE PLANTATION SDN. BHD.	LOT 1223, PARIT TURUN,	LEDANG,	BUKIT GAMBIR	JOHOR	102.6546	2.211	01020056
FONG SOON POULTRY FARM SDN. BHD.	5 GROUND FLOOR, JLN CENGAL	OFF JALAN GENUANG, SEGAMAT	SEGAMAT	JOHOR	102.82754	2.500400004	01024045
INN SOON FARMING SDN. BHD.	NO.188, JALAN ISMAIL,	SENGGARANG,	BATU PAHAT	JOHOR	103.0592733	1.74529556	01029039
BBC PELITA PLANTATION (JEPAK) SDN. BHD.	NO.84 KAMPONG GELANG CHINCHIN	SEGAMAT	JOHOR	JOHOR	102.8142546	2.573063058	01144049
LEW PENG ENTERPRISE SDN. BHD.	LOT 1040 JALAN KAMPUNG MURNI JAYA	LAYANG - LAYANG	KULAI. JOHOR	JOHOR	103.5379129	1.80530001	01134047
FISH WORLD AQUARIUM TRADING	LOT 207, LORONG SEROJA 2,	TAMAN MEDAN JAYA	KOTA TINGGI	JOHOR	103.8834953	1.722812982	01135018
LADANG FELDA INAS SELATAN	LADANG FELDA INAS SELATAN	MUKIM SEDENAK	KULAI	JOHOR	103.4452862	1.726078126	01007013
SLK GEMILANG SDN. BHD.	NO. 37 JALAN NAGASARI	SEGAMAT BARU	SEGAMAT	JOHOR	102.85254	2.493529987	01142017

Step 2: Upload the excel file into ARCGIS and ARCGIS Enterprise

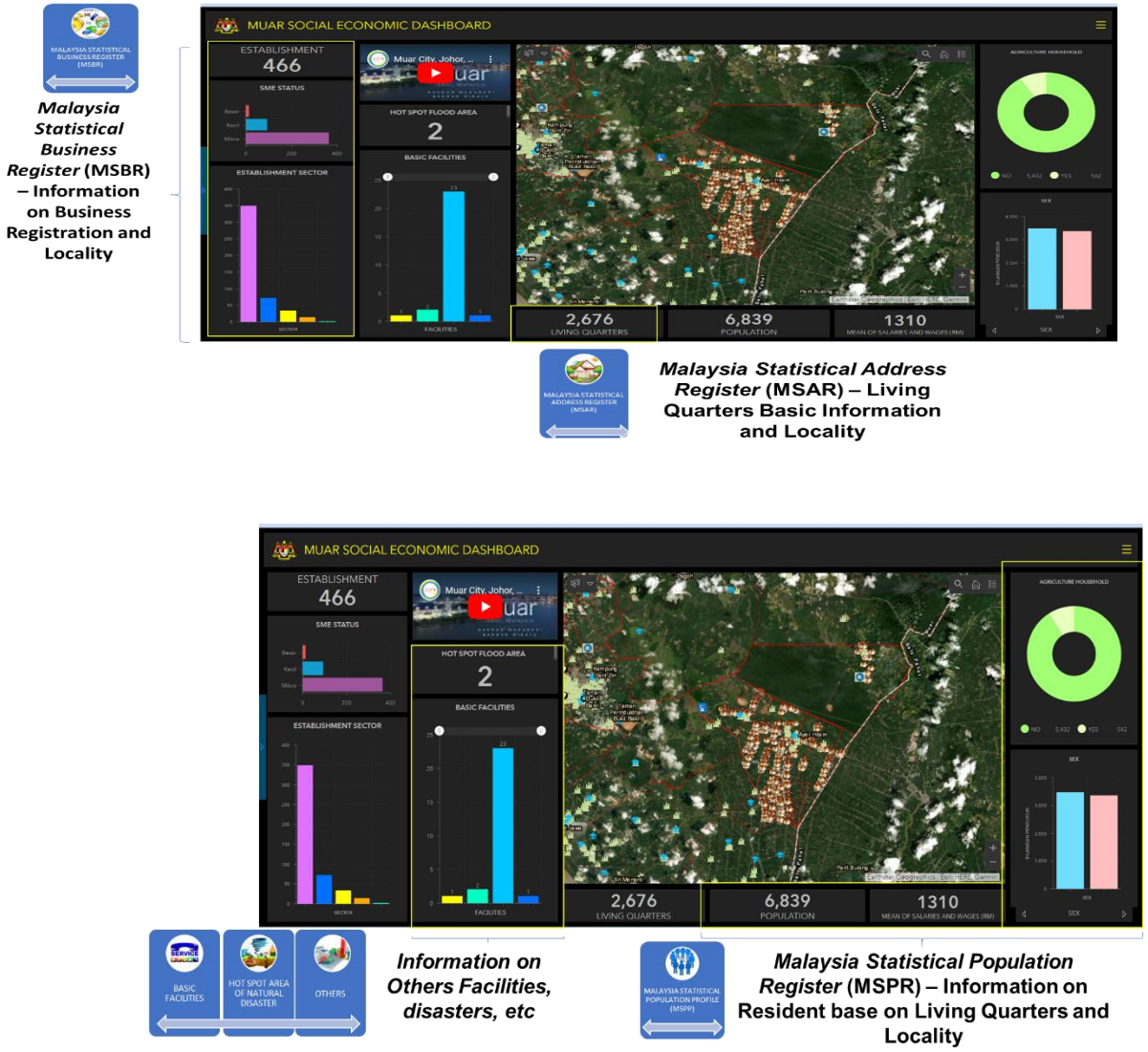
Excel file with the selected variables, such as establishment, living quarters and other basic facilities (schools, homestays etc.) will be compiled in one file (zip. file).







Step 3: Visualisation of linkages MSBR, MSAR and MSPP by using ARCGIS Enterprise



The case study indicated that the location of each establishment is specified using map coordinates, giving the exact location of each establishment as well as the living quarters and basic facilities.

Table 1: Number of establishments, living quarters and population by enumeration blocks in Mukim Air Hitam, Muar, Johor, Malaysia.

EB	No. of Establishment	No. of LQ	No. of population
01138001	18	119	347
01138002	13	146	392
01138003	14	116	322
01138004	7	106	222
01138005	4	81	203
01138006	12	146	400
01138007	15	138	310
01138008	2	65	178
01138009	0	73	215
01138010	5	134	322
01138011	5	104	330
01138012	8	137	434
01138013	3	95	284
01138014	7	93	291
01138015	2	130	404
01138016	1	67	219
01138017	0	78	279
01138018	6	121	359
01138019	8	105	282
01138020	7	129	444
01138021	0	104	337
01138022	3	77	224
Total	150	2,364	**6,798**

Benefits of integration and linkages business and social statistics with spatial data

The integration and linkages businesses and social statistics with spatial data are very important to provide a solid foundation for reliable and accurate statistical area. It can be used to identify the target establishment and population especially during a pandemic or disaster. If there is any pandemic for example, during the previous COVID 19, most businesses were affected especially the micro, small, and medium establishments. Thus, if this database is available, the assistance or aid delivery for those affected can be more efficient. It can be stratified by the size of the establishments (micro, small, and medium). In terms of the population, updating the population profile for each living quarters enables assistance or aid delivery be better targeted and prioritised. For example, different assistance or aid is needed for the elderly, children, unemployed, disabled and others.

Any planning and implementation can be implemented efficiently due to the readiness of data down to even small areas. Different localities have different planning and implementation. Local authorities, politicians and researchers can use the small area statistics for the planning and implementation by targeted area or locality.

In addition, the monitoring for birth and death establishments as well as population migration will be more easily carried out with the updating of business and social statistics with spatial data. Latest profile for establishments and population can also be one of the input for inter census of Malaysia Population and Housing census, thus, reducing the operation cost, financial and human capital required.

This case study acts as a push factor for local authority and create awareness on the latest and up to date profiles of establishments and population at local area. It is a must in order to provide efficient aid to targeted population (establishments and people).

Challenges

There are some challenges to integrate and create linkages the MSBR, MSAR, MSPP with spatial data particulary in the aspect of collaboration, data quality and human capital. In order to have up to date profiles of the establishments and population of localities, a strong and good collaboration with the district local authorities and communities is essential. The awareness and willingness of community to self-update their latest profile is essential since the census is conducted once in 10 years, while surveys just comprise selected samples. In terms of administrative data, there are also limitation whereby the data was compiled based on different objectives for the different sources. For example, the main source of establishments list is from CCM, which is dedicated to registered companies. Meanwhile, whether is the companies are really operating on the ground needs to be validated especially those companies without financial reports, such as micro and small companies.

Data quality is also a major challenge. The changes of establishments and population are quite dynamic. Different sources have different objectives. Hence, the definition and metadata are also different. Some sources do not use the business registration number, merely based on registered company names, while the birth and death of a company may also not be registered or updated in CCM. Again, collaboration and compilation from the district level is a must to ensure this case study succeed and maintained.

In terms of human resources, is very limited since this case study was based on existing resources. Engagements need to be done at many levels, such as at district level then up to village or mukim level to justify to the community. In this situation, collaborations among district offices, head of communities or villages, DOSM and state DOSMs is pertinent. Then, to do the integration and linkages from many databases with spatial data and visualisation through dashboard, also needs specific skilled officers.

Way forward.

This case study has been extended to other districts or localities in other states in Malaysia, such as Besut (Terengganu), Langkawi (Kedah), Perak (Perak Tengah), Seremban (Negeri Sembilan), Kota Bharu (Kelantan) and Pekan (Pahang). Meanwhile, for the Agriculture Census 2024 preparation, about 20,000 establishments were geocoded based on their addresses using the Geocode by Awesome Table. However, it needs to be verified by state DOSM. Moving forward, all the establishments in MSBR will have been geocoded based on their addresses and will be allocated to state DOSM for verification.

Conclusion

The first case study in Mukim Ayer Hitam, Muar, Johor is undertaken by the DOSM in collaboration with the District Office of Muar, Johor. This case study was initiated with the aim to inspire the other districts in Malaysia. The success of this project depends a lot on the strong collaboration between DOSM headquarters, state DOSM, District Offices, heads of villages / communities and communities. One of the main outcomes of this project is to produce a solid foundation for reliable and quality statistics based on locality. Hence, any policy and aids at the district level can be targeted based on the specific needs and groups.

Acknowledgement

The authors would like to acknowledge the Population and Demographic Statistics Division, Department of Statistics Malaysia and States office of Johor, Department of Statistics Malaysia for their contribution to this paper.

.

REFERENCES

Ryan, L., Thompson, C. And Jones, J. (2020) 'A Statistical Business Register Spine As A New Approach To Support Data Integration And Firm-Level Data Linking: An ABS Perspective', *Statistical Journal Of The IAOS*, 36(3), Pp. 767–774. Doi:10.3233/Sji-200640.

Eileen, C. and Irene, T. (2017) Establishing the Statistical Business Registers in Asia.

Register-based statistics in the Nordic countries (2007). New York, NY: United Nations.

Population Register data, basis for the Netherlands Population Statistics (2017). Statistics Netherlands, CBS.

Mariana, K. Et Al. (2019) Implementing A Geospatial Data Strategy In The European Statistical System [Preprint].

United Nations Guidelines on Statistical Business Register (2020)

Siti Haslinda, M.D. and Nur Aziha, M. (2016) Malaysia Statistical Business Register (MSBR) Innovates New Insight of Malaysia Trade Statistics.

ABS Address Register, Users' Guide | Australian Bureau of Statistics

Tey, N. Et Al. (2021) 'Spatial Demographic Data For Planning And Research', PLANNING MALAYSIA, 19. Doi:10.21837/Pm.V19i15.927.

Zainal Abidin, A.M., and Muhamad Fadzil, I., Nadia, M., (2020) Spatial Analysis: Ageing Population of Multi-ethnic in Rural Area Malaysia. Available at: https://www.unescap.org/sites/default/files/70_Spatial_Analysis Ageing Population of Multi-ethnic in Rural Area Malaysia.pdf