

United Nations Statistical Commission

Fifty-second session

1–3 and 5 March 2021

Item 3 (j) of the provisional agenda

Items for discussion and decision: big data

Document E/CN.3/2021/14 – Report of the Global Working Group on Big Data for Official Statistics

Statement provided by:

Dr. Mohd Uzir Mahidin, Chief Statistician,
Department of Statistics, Malaysia (DOSM)

Statement:

Malaysia highly appreciate the effort of the Committee of the Global Working Group on Big Data for Official Statistics. The set-up of UN Global Platform and the Regional Hubs as a collaborative environment is a good effort to explore ideas constructively and able to look beyond than NSO's own limited vision. When a variety of knowledge and skills are pooled, it creates a talent pool that is vast, competent, able and experienced. Virtual collaboration minimizes or even eliminates the need for finances.

The in-kind contributions from global statistical community experts provide plenty of opportunities for gaining new skills or a unique set of skills and experience as well as for know-how exchange with other colleagues.

The encouragement and support of a global program on training and certification in the use of Big Data and data science for official statistics are a great way to gain an edge because certifications allow statistician to develop skills and at the same time it is a way to validate the possessed skills.

DOSM's officers have been trained with a variety of SAS tools and open source tools to gain insights from big data. Nine (9) DOSM officers have been recognized as SAS Certified Data Scientist in 2018 and DOSM also encourages its officers to further studies in Data Science field. In May 2020, Big Data Analytics Core Team has been formed by gathering DOSM's officers who has certification in Data Science and those who possesses the related skills. DOSM will keep leveraging internal skills to bring the basic data science expertise in house for advanced analytics. Among key data science area to be developed are data engineering, statistical science, mathematics, visualization, machine learning, data mining, data modelling and data visualization.

Apart from hiring new talents to strengthen institutional capacities, to support the demand of the emerging skills, DOSM also encourages existing capacities to have multi-talents and equip themselves with in-demand skills through continuous learning in any platforms specifically via online training, workshops and conference. This includes compulsory tailor-made training outlined by the Department and also relevant platforms offered by other public and private agencies.

To ensure DOSM capacities are in line with the in-demand emerging skill sets and resources in this data revolution era, DOSM has planned strategic collaborations with few government agencies as well as private sectors to implement big data-related projects.

On top of that, DOSM has requested to participate in the ONS-UNECE Group for Machine Learning 2021 (ML 2021) project. With the aims of developing research, building skills and sharing resources on Machine Learning developments and applications for official statistics across the global statistical community, DOSM believes this group can provide the platform for the team to accelerate the

United Nations Statistical Commission

Fifty-second session

1–3 and 5 March 2021

Item 3 (j) of the provisional agenda

Items for discussion and decision: big data

learning curve. The Estimating of Malaysia Rubber Plantation Area and Productivity Using Satellite Imagery and Machine Learning has been proposed under this ML 2021 project. The objective is to identify new and potential approach to complement and/or or replace the traditional data collection method that is presently applicable in rubber production monitoring and estimation. It is expected to provide a swift and speedier estimation of rubber production at any point of time.

Dealing with various global restrictions related to the COVID-19 pandemic, DOSM like other NSOs is adapting to the ongoing challenges as the pandemic continues to keep affecting the operations. Thus, DOSM is seeking to increase the role of administrative data in the production of official statistics especially information on geographical references. A complete and accurate address information is critical as many of the statistical outputs rely heavily on it. Multiple sources of dataset from multiple government agencies keep address data in variety format and address matching process becomes a complicated task. With recent innovations in machine learning, there is a potential to apply machine learning to the address matching process. Address Index (AI) project done by ONS could be one of the references for DOSM to pursue with the project.

In addition, DOSM also increase the usage of administrative data and reduce respondent burden by collaborating with Ministry of Domestic Trade and Consumer Affairs Malaysia in order to strengthen the compilation of the Consumer Price Index. At the same time, DOSM is leveraging on the web crawling and web scraping by selecting significant website from the new online data sources. In order to improve this project, DOSM would like to collaborate with any NSO in this field.

Besides, brings international experts and local talent together to collaborate in big data project would be another way for DOSM to build a workplace culture centered around learning and development. Any issues can be encountered by asking for feedback and opinions, debating for potential solutions and gaining a better sense of how experts work.

Making connections with global statistical community experts, **DOSM also willing to share experiences on the use of big data specifically in the compilation of informal sector statistics, Trade by Enterprise Characteristic (TEC), and Calculation method of Malaysia's New Poverty Line Income and Multidimensional Poverty Index (MPI).**

Submitted on:

2/17/2021