Delighting in DATA

A treasure trove of data is now at everyone's fingertips, allowing anyone to gain deeper insight into the country.

By CHRISTOPHER FAM
lifestyletech@thestar.com.my

DID you know that a 10kg bag of rice is cheapest in Perlis and Pahang, at only RM22.88, while the very same bag can fetch as much as RM30 in Sarawak?

These and other interesting statistics can be found on the NextGen OpenDOSM portal created by the Department of Statistics Malaysia (DOSM).

The platform has been making waves on social media since it got a major revamp early this month, with both experts and enthusiasts praising it for its potential uses.

And in just five days of its "upgrade", over 33,000 users had accessed the platform and information dashboards, as well as downloaded raw data from the site.

But the big question is: How does the average Malaysian benefit from the vast trove of data?

In-depth information

On the website, OpenDOSM is described as: "A platform that catalogues, visualises, and analyses Malaysia's wealth of data.

"Everything on this site is open-sourced and freely available for the nation's benefit. If data is the new oil, then openness is the pipeline that maximises its value."

It currently hosts 168 different datasets divided into 12 themes with 16 information dashboards.

Examples of thematic data include petrol pricing and economic indicators like the Consumer Price Index (CPI), exchange rates, gross domestic product (GDP), and employment rates.

It also hosts pricing data from the Domestic Trade and Cost of Living Ministry's *PriceCatcher* app, which boasts 28 million entries, including the price of rice cited earlier.

The datasets even include amusing tidbits such as the increase in demand for red notes during Chinese New Year and green notes during Hari Raya Aidilfitri.

According to the chief statistician, Datuk Seri Dr Mohd Uzir Mahidin, the goal of OpenDOSM is to provide data to the public in an open and seamless manner with minimal downtime.

"Datasets in this platform are presented in a data catalogue and dashboard with comprehensive metadata. It also contains the limitations of use for each dataset.

"OpenDOSM benefits the public in such a way that datasets are summarised and presented in easy-to-understand graphs and tables, which can be used as a reference or input for informed decision-making," he said.

Charting
a new course

A big reason behind the creation of the platform was to bridge the knowledge gap between the country's leaders and the public by ensuring that everyone has access to the same information.

"Knowing the areas of concern helps Malaysians gain insight into the reasoning behind the budget allocations towards certain areas.

"For instance, if the public knew about the crime rate of a certain area, they would understand why the government would allocate the necessary resources to combat the issue in the area.

"But if they notice that the allocation doesn't match the data, the opportunity arises to question whether it is an efficient use of public funds, and have their point backed up by data.

"This could cover blind spots that were missed at the macro level, as it can be hard to tell what's wrong with individual trees when looking at an entire forest," he said.

Mohd Uzir also brought up the development of a "data-centric" culture among Malaysians as the ultimate goal of OpenDOSM and hopes to achieve two things along the way.

"The goal is, first and foremost, to improve public perception and awareness of data-driven decision-making while also contributing to new 'value creation'.

"And secondly, to encourage other government agencies to collaborate in data sharing and make it available to the public, while preserving the integrity, provenance and attribution," Mohd Uzir said.

Earlier this month, he invited other government departments to share their data on the platform to make it more vibrant.

He added that it is only a matter of time until private organisations and companies adopt the initiative to share their data with the public, and OpenDOSM will be ready when it happens.

According to the "release calendar" on the DOSM website, not only will more datasets be made available to the public, but existing ones will be updated regularly.

Examples given by Mohd Uzir of the types of data that could be made available include the likelihood of employment based on the area of study and industry, average salaries and tourism hotspots.

"We want to make data a part of our culture, for people to make it the topic of conversation at dinner or over coffee.

"Perhaps data saying that a friend's hometown has a high number of schools or that the state you come from has high rainfall could become staple topics of conversation with the data being available openly," he said.

Simplifying statistics

Subject matter experts, such as academic researchers, are likely to benefit the most from free access to demographics, crime rates, and raw economic data but steps are being taken to ensure that the data will be useful to all.

"Using football as an analogy, a professional player would fare much better than an amateur simply because of the depth of his knowledge and skills.

"The same can be said for data, so those with specialised skills are in a position to guide the public on how it should be done," Mohd Uzir said.

The long-term goal of OpenDOSM is to help the country build a 'data-centric' culture, according to Mohd Uzir. — Photos: AZHAR MAHFOF/ The Star

"The launch of OpenDOSM was part of the effort to make data more understandable to the public – the original DOSM website was not as easily readable.

"But at the end of the day, there are still skills required to analyse and interpret data.

"Using football as an analogy, a professional player would fare much better than an amateur simply because of the depth of his knowledge and skills.

Some have even gone so far as to create customised dashboards based on their interests or expertise, while others have made tutorials on using the platform.

DOSM has also conducted webinars on using the platform, which will continue if there's a demand for them, and has started looking into outreach programmes in

He added that this has already been happening on Twitter, with more advanced users sharing how they are using the data, while exchanging ideas and tips with the community.

schools at the state level.

"For some countries, even cab drivers at the airport know and can talk about the socioeconomic situation of their country alongside tourist spots.

"In a way, having this knowledge about your home country can serve to promote the country to foreigners and encourage tourism.

"But first we need to change the landscape of our perception of statistics," Mohd Uzir said.

Guidance or misguidance?

While relying on experts will help illuminate the way, it will do no good if the data is instead used to mislead the public or make questionable claims.

Those with malicious intent could use the data that is favourable to themselves or their argument, leaving out the bigger picture while appearing legitimate because of their source.

"There is always the possibility of data misinterpretations, be they intended or unintended.

"To avoid such occurrences, this platform provides comprehensive metadata, data explanations, and

limitations for each dataset for users to understand and be cautious in data interpretation.

"Nevertheless, intended misinterpretation is unavoidable if users themselves understand the data context well but purposely misuse it," Mohd Uzir said.

He added that, in the end, the biggest protection against it is having literacy in reading data, coupled with the openness of social media.

Although there is no way to stop it, involving others in the discourse can help expose unscrupulous actors and let the facts speak for themselves, he added.

To that end, the OpenDOSM platform maintains a level of transparency, providing metadata, methodology, caveats, and even source code to help things along.

Figuring out the future

As the library of data continues to expand over time, the sky's the limit on the potential uses for the public repository.

However, those who want to harness the data will have to equip themselves with the right analytical tools for deeper insights.

These insights are expected to give everyone a better understanding of the country, helping them think of how they can help change it for the better moving forward.

"We've been commended by various international statistical institutions for our efforts to align with open data standards.

"In the long term, we hope to see the development of critical thinking and creativity built on top of our data foundation.

"It could even serve as a learning tool, but for that to happen, it will require literacy, which will only develop over time.

"In the future, we hope to bring granular data that goes from the top, at the national level, all the way down to towns and villages to the platform.

"This will be on top of bringing other features like 'presentation creation' that will be made available to the public as well," said Mohd Uzir.