



## OPEN DATA PLATFORMS

Prepared by: **Roziana Shamsuri & Liza Ab Llah**  
Research and Information Services Division,  
Perpustakaan Sultan Abdul Samad, Universiti Putra Malaysia

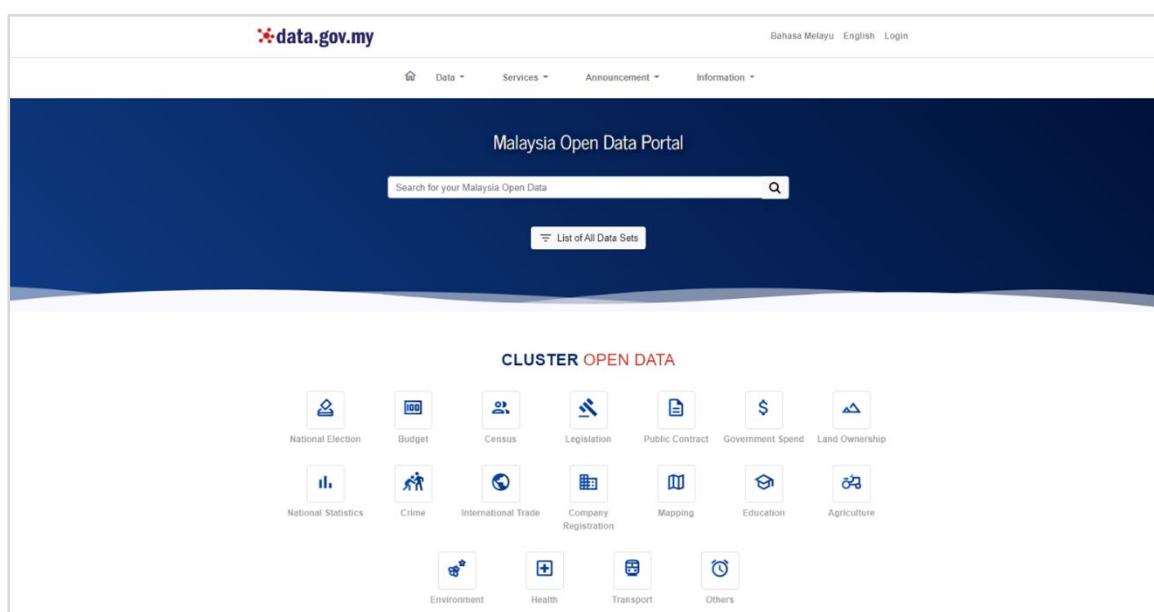
### ***What is Open Data?***

Open data is information or content made freely available to use and redistribute, subject only to the requirement to attribute it to the source. The term also may be used more casually to describe any data that is shared outside the organization and beyond its original intended use, for example, with business partners, customers or industry associations. (Gartner Information Technology Glossary, 2014)

If you are a data enthusiast and need datasets for practice data exercises, here are some of the open data portals available. You can use the data provided for free too. Let's check it out!!

#### 1. [Malaysia Open Data Portal](#)

The portal is managed by the Malaysian Administrative Modernization and Management Planning Unit (MAMPU). The Malaysia Open Data Portal serves as the heart of the country's open government data initiative. The site collects datasets from about 390 dataset suppliers, from the various ministries to state governments, police authorities and even the Department of Statistics Malaysia, housing about 12,700 datasets. The language either in English or Bahasa Malaysia.



## 2. [National Property Information Centre](#)

The portal contains large databases on the local property market such as the status of newly launched residential properties, nationwide construction activities and total property supply. Most of these databases are not in their rawest format, featuring interpreted results and processed reports, thus making it difficult to repurpose them for machine-learning purposes. However, these reports are still useful in generating insights and conducting basic analysis. The portal additionally gives unpublished data on a pay-per-use basis. The charges are not publicly available, but as a reference, the annual property market report itself is priced at RM100. Currently, these unpublished databases are available only to nationally recognised and registered property valuers and real estate agents, and it is uncertain whether the service will be available to the general public.

## 3. [Malaysia Open Data Ocean](#)

openDOSM is an alternative platform for open data on official statistics produced by the Department of Statistics, Malaysia. This is an initiative to expand the open data audience and to promote open data usage among public. Datasets covers wide range of topics, including, social, economic & financial and environment. Data selection is based on availability, real-time, near real-time, and the latest administrative data. The portal also offers publications or detailed unpublished data from DOSM. Requests for data will be charged on cost recovery basis, depending on the amount of resources used to process and compile the data.

#### 4. [Kaggle](#)

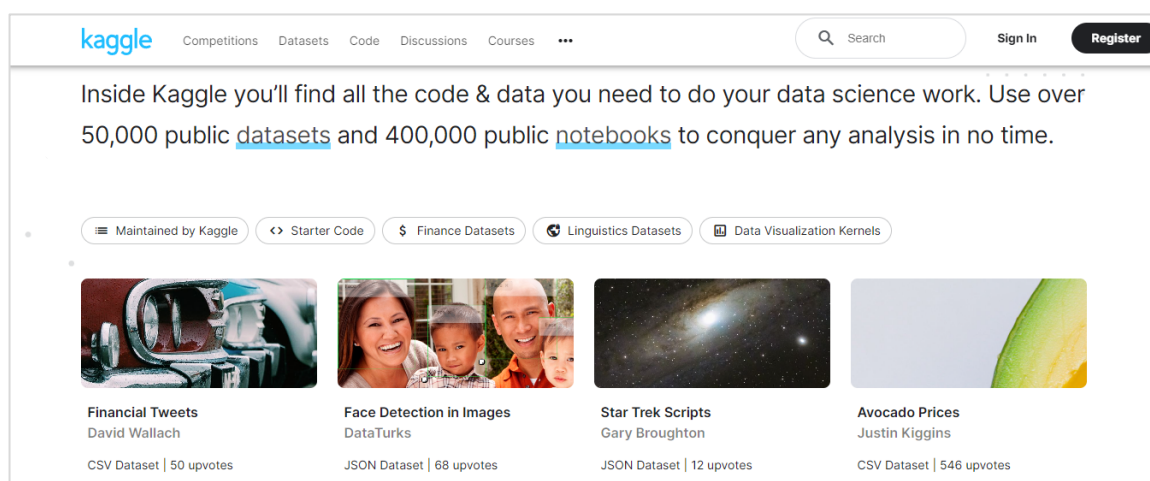
Kaggle is an open data international company that offers its data to the public. Kaggle offers a no-setup, customizable, Jupyter Notebooks environment. Get access GPUs at no cost to you and a huge repository of community published data & code. Inside Kaggle you'll find all the code & data you need to do your data science work. Use over 50,000 public datasets and 400,000 public notebooks to conquer any analysis in no time.

In the competition, grow your data science skills by competing in our exciting competitions. Find help in the documentation or learn about Community Competitions. To make learning fun, use Kaggle with no-cost, self-service platform to create competitions for your community. Learn more or create a competition to be considered for a \$5,000 monthly award. Kaggle's inClass competition programme is also free for universities and is being used by institutions such as Harvard, Stanford and Oxford Universities. Students can even use existing codes provided by the Kaggle community and repurpose them for their personal projects.

In the dataset, you can explore, analyze, and share quality data. data types, creating, and collaborating. They have trending datasets, popular datasets, music, business, computer science, classification and standardized testing. Each dataset can search by filter the categories.

In the discussion part, it will discuss the Kaggle platform & machine learning topics. This includes sharing feedback, asking questions, and more.

You also can gain the skills to do independent data science projects. We pare down complex topics to their key practical components, so you gain usable skills in a few hours (instead of weeks or months). The courses are provided at no cost to you, and you can now earn certificates.



#### 5. [OECD iLibrary](#)

The Organization for Economic Co-operation and Development (OECD) is an international organization that works to build better policies for better lives. Our goal is to shape policies that foster prosperity, equality, opportunity and well-being for all.

Working hand in hand with policy makers, governments and citizens, the OECD informs public decision-making bringing together 38 member countries on key global issues.

Datasets from the Organisation for Economic Co-operation and Development (OECD) is a must-have for any macroeconomic data analysis. OECD's iLibrary contains books, papers and statistics on the economic conditions of all 38 of its member countries, and its content is widely used by universities, research organizations, think tanks and public administrators.

In addition to economic indicators, our site also contains datasets ranging from energy to education, with content released by the International Energy Agency, Nuclear Energy Agency, Programme for International Student Assessment, and International Transport Forum.

Many of the datasets go as far back as 1998, giving data scientists plenty of wiggle room to conduct any time-series or historical analysis. Even without the raw datasets on the site, the OECD publishes between 300 and 500 books annually — a repository of information useful for any knowledge analysis beyond pure data and should be part of any data scientist’s toolkit.

The screenshot shows the OECD iLibrary website interface. At the top, there is a search bar with the text 'Search all content by title or author' and a magnifying glass icon. To the right of the search bar are links for 'EN', 'My Favorites', and 'Logout'. Below the search bar is a navigation menu with options: 'Browse by Theme', 'Browse by Country', 'Browse by Theme and Country', 'Catalogue', 'Statistics', and 'Podcasts'. The main content area features a 'What's New?' section. On the left is a book cover for 'Building Trust to Reinforce Democracy: Main Findings from the 2021 OECD Survey on Drivers of Trust in Public Institutions'. To the right of the cover, the text reads: 'What's New? Building Trust to Reinforce Democracy Main Findings from the 2021 OECD Survey on Drivers of Trust in Public Institutions'. Below this, a short description states: 'This report presents the main findings of the first OECD cross-national survey on trust in government and public institutions, across five drivers of trust – reliability, responsiveness, integrity, openness, and fairness – and provides insights for future policy reforms.'

## 6. [Asia Open Data Portal](#)

Dataportal.asia is one-stop data access services and aims at facilitating and providing advanced features for cross-border open data access to ensure the exploration, exploitation, and re-usability of open data across Asian countries.

The Asia Open Data Portal contains about 167,000 datasets. It is Asia’s first official open data portal, set up by the Taipei Computer Association. Dataportal.asia harvests the metadata of public data made available across Asia. The metadata contains the information made available with the data set by the initial publisher that offers different and multiple ways for users to access the data set they need instantly such as by countries, data format and themes.

The catalogs of datasets within the portal can be massive but are neatly organized by country and broad categories, such as science and technology, population and society, labor and health.

It has sourced almost all of its Malaysian datasets from the Malaysian open data portal. The browsing experience is nearly identical to visiting the official portal itself. Still, it provides the added advantage of quickly accessing datasets from around Asia without having to open multiple browser tabs.

The screenshot shows the Dataportal.asia website. The header includes the logo 'Dataportal.asia' with the tagline 'Data explorer in Asia' and a navigation menu with links: 'Member', 'Datasets', 'Catalogues', 'Categories', 'Applications', 'Links', and 'About'. The main content area is divided into two columns. The left column features a 'Welcome to Dataportal.asia' section with a map of Asia titled 'Data Explorer in Asia'. Below the map, a short description states: 'Dataportal.asia aims at facilitating and providing advanced features for cross-border open data access to ensure the exploration, exploitation, and reusability'. The right column displays search statistics: 'Served 17,851,714 Attendances', '167.3k datasets', '21 catalogues', and '12 categories'. Below the statistics are two search boxes: 'Dataset Search' with a search bar containing 'e.g., environment' and a magnifying glass icon, and 'InSite Search' with a search bar containing 'Not limited to searching datasets. e.g., COVID-19' and a magnifying glass icon.

## 7. [World Bank Open Data](#)

The World Bank recognizes that transparency and accountability are essential to the development process and central to achieving the Bank's mission to alleviate poverty. The Bank's commitment to openness is also driven by a desire to foster public ownership, partnership and participation in development from a wide range of stakeholders. As a knowledge institution, the World Bank's first step is to share its knowledge freely and openly.

The World Bank Open Data site is designed to make data easy to find, download, and use. All of the data found here can be used free of charge with minimal restrictions. A comprehensive listing of data and datasets published by the World Bank is available through the Data Catalog. The Catalog contains all the datasets in DataBank, plus many other useful datasets, including some sub-national and raw data from surveys. Datasets can be easily downloaded, or accessed through the DataBank query tool or custom tools specific to the dataset. The World Bank contains about 5379 datasets. It is a valuable tool to support research by journalists, academia and others, broadening understanding of global issues.

**THE WORLD BANK** | Data Catalog

HOME DATA COLLECTIONS GETTING STARTED FAQS LOGIN

**GETTING STARTED**  
The Data Catalog is designed to make World Bank's development data easy to find, download, use, and share. It includes data from the World Bank's microdata, finances and energy data platforms, as well as datasets from the open data catalog... [READ MORE](#)

**Featured**

- FINANCE FOR AN EQUITABLE RECOVERY**  
**World Development Report 2022**  
The World Development Report 2022: Finance for an Equitable Recovery examines the central role of finance in the economic recovery from the COVID-19 pandemic. In doing so, it looks into the consequences of the crisis most likely...
- Statistical Performance Indicators**  
National statistical systems are facing significant challenges. These challenges arise from increasing demands for high quality and trustworthy data to guide decision making, coupled with the rapidly changing landscape of the data revolution.
- World Development Indicators**  
The primary World Bank collection of development indicators, compiled from officially-recognized international sources. It presents the most current and accurate global development data available, and includes national, regional and global estimates.
- COVID-19 Data**  
Data relevant to the coronavirus pandemic, drawn from the World Bank's data catalog and other authoritative sources.

### Sources:

Dataportal.asia. Retrieved June 9, 2022, from <https://dataportal.asia/home>

Gartner. *Gartner Glossary: Open Data*. Retrieved April 5, 2022, from <https://www.gartner.com/en/information-technology/glossary/open-data>

Kaggle. Retrieved June 9, 2022, from <https://www.kaggle.com/>

Malaysia Open Data Portal. Retrieved May 27, 2022, from <https://www.data.gov.my/>

OECD iLibrary. Retrieved June 9, 2022, from <https://www.oecd-ilibrary.org/>

OpenDOSM. Retrieved May 27, 2022, from <https://open.dosm.gov.my/>

Pusat Maklumat Harta Tanah Negara. Retrieved May 27, 2022, from <https://napic.jpph.gov.my/portal/web/guest/main-page>

The Edge Market. *Listicle: Open data platforms every Malaysian data scientist should know*. Retrieved April 5, 2022, from <https://www.theedgemarkets.com/article/listicle-open-data-platforms-every-malaysian-data-scientist-should-know>

The World Bank. Retrieved June 9, 2022, from <https://datacatalog.worldbank.org/home>