

Google 25 million free datasets



Harman Bhutani

Jan 10 · 2 min read ★

Dataset Search

Search for Datasets



Try [coronavirus covid-19](#) or [education outcomes site:data.gov](#).

[Learn more](#) about Dataset Search.

Datasetsearch, a free tool for searching for 25 million publicly accessible datasets, was recently published by Google.

The screenshot shows the Google Dataset Search interface. At the top, the Google logo is on the left, and a search bar contains the word "Covid". To the right of the search bar are icons for help, a list, and a user profile. Below the search bar, there are filters: "Last updated", "Download format", "Usage rights", "Topic", and "Free". A "Saved datasets" button is on the right. The search results section shows "100+ datasets found". The first result is from the World Health Organization (WHO) titled "WHO Coronavirus disease (COVID-19) situation reports" with a PDF icon and a link to "Explore at www.who.int". Below this is a result from the CDC titled "Coronavirus Disease 2019 (COVID-19)" with a link to "www.cdc.gov". The third result is from The New York Times titled "Coronavirus (Covid-19) Data in the United States" with a link to "www.nytimes.com" and "github.com". On the right side of the WHO result, there is a detailed view showing the WHO logo, the title "WHO Coronavirus disease (COVID-19) situation reports", a blue button "Explore at www.who.int", a PDF icon, and sections for "Dataset provided by" (World Health Organization), "Area covered" (Global), and "Description" (Daily situation updates and data regarding the COVID-19 outbreak). The description includes two bullet points: "Figure 1: Countries, territories or areas with reported confirmed cases of COVID-19." and "Table 1: Confirmed and suspected cases of COVID-19 acute respiratory disease reported by provinces, regions and cities in China."

The search tool includes filters to limit results based on their license (free or paid), format (CSV, images, etc), and update time.

Note-Google does not curate or provide direct access to the 25 million datasets directly.

Google depends on publishers of datasets to use schema.org's open standards to define the metadata of their datasets. Then Google indexes the metadata and makes it searchable through publishers.

Since publishers are still expected to host the datasets themselves, Google will also have their datasets indexed by for-profit publishers that adhere to schema.org standards. I found that half of the datasets in the search results in my anecdotal experience came from for-profit aggregators, with an even higher percentage when looking for market-related datasets.

To publish your own datasets, you can simply use the open-standards of schema.org. The number of publicly available datasets is likely to continue growing as more publishers conform to the standard.

At this time, Google does not provide an API for searching or downloading the free datasets.

You can find more information on the same at [Google Dataset Blog](#).

Thanks for reading! I write on Medium about data science-related topics, so you can follow me to get notified when I write more articles like this one.

And don't forget to press the clap button if you liked the article :)

Follow me on [LinkedIn](#) if you'd like to stay in touch!

Sign up for Analytics Vidhya News Bytes

By Analytics Vidhya

Latest news from Analytics Vidhya on our Hackathons and some of our best articles! [Take a look](#)

Get this newsletter

Emails will be sent to byteharman0101@gmail.com.
[Not you?](#)

Data Science

Machine Learning

Python

Artificial Intelligence

Google

About Help Legal

Get the Medium app

