

## **A review of the resource Google Dataset Search**

Katherine Howell

Online Learning and Information Literacy Librarian and Business Liaison, Raymond Cravens Library, Western Kentucky University, [Katherine.Howell@wku.edu](mailto:Katherine.Howell@wku.edu)

Google. (n.d-a.). *Dataset Search*. <https://datasetsearch.research.google.com/>

I have found Google's Dataset Search tool to be helpful regardless of discipline, but it is especially helpful for performing marketing and economics research. It is a search engine that searches metadata for datasets hosted in thousands of repositories across the Web. The goal of the tool is to make datasets more universally accessible and useful. In addition, the tool's creators hope to create a data sharing ecosystem that will encourage publishers to follow best practices for data storage and to give the creators of datasets a way to show the impact of their work through citations. According to Google, "as more dataset repositories use schema.org and similar standards to describe their datasets, the variety and coverage of datasets that users find in Dataset Search will continue to grow" (Google, n.d.-b). As of 2020, when the tool came out of beta, the Dataset Search corpus contained more than 31 million datasets from more than 4,600 internet domains (Noy & Benjelloun, 2020).

The largest subject area covered by the corpus is Social Science at 26.2%, which makes it particularly useful for finding demographic information for secondary marketing research (Noy & Benjelloun, 2020). Users can search for datasets on Dataset Search using a simple keyword search. I recommend using keywords that are as specific as possible, as there are limited options for filtering results: results can be filtered by date last updated, download format, usage rights, broad topic, and whether access is free. So, for example, if a student is performing marketing research and their target market is elderly U.S. adults, they could search for "population elderly U.S. adults." That search string returns over 100 datasets, from repositories like Statista, Federal Reserve Economic Data (FRED), Figshare, and the Inter-university Consortium for Political and Social Research (ICPSR).

To learn more about the datasets, which appear in a list on the left side of the webpage, users can click on a search result to get more information. Each dataset entry includes a wealth of metadata, including but not limited to a unique identifier, date uploaded, provider/repository, authors, license for use, time period covered, area covered, funding source, and a description. At the top of each entry, Dataset Search provides the available download format of the data. Once a user has found a dataset that addresses their topic and is in the format they require, they can simply visit the blue "Explore" link, where they will be taken to a repository's site for download.

One especially helpful characteristic of the tool is that it displays the number of scholarly articles citing the dataset according to Google Scholar data. For example, the dataset "Working Age Population: Aged 15-64: All Persons for the United States" from FRED has been cited by 24 scholarly articles (Google, n.d.-a). I place special emphasis on this feature when teaching instruction sessions, as it can help users evaluate the authority of the source. I tell students that if many scholars have used the data, it is most likely reputable and a good choice for their research.

K. Howell

In conclusion, Google's Dataset Search is a unique resource that allows users to search for datasets across the Web using a simple keyword search. It is easy to filter results to only those that are free and able to be reused. By utilizing the tool, users can become familiar with proper data management techniques, which may encourage them to share data of their own in the future.

## References

Google. (n.d.-a). *Dataset Search*. <https://datasetsearch.research.google.com/>

Google. (n.d.-b). *User Support Center*. <https://datasetsearch.research.google.com/help>

Noy, N., & Benjelloun, O. (2020, August 25). *An Analysis of Online Datasets Using Dataset Search (Published, in Part, as a Dataset)*. Google AI Blog.  
<https://ai.googleblog.com/2020/08/an-analysis-of-online-datasets-using.html>