

Finding & Getting Data - Google Dataset Search

Another source for data using Google tools is a tool called "Google Dataset Search". This tool is actually an aggregator of data sources. So imagine using Google to look for sites in pages that publish only data, but actually from official sources. So, this tool tries to look at a plethora of websites and then it gets the information, the metadata, from these websites to talk about the data that they publish.

So let's, let's type in in the Google Dataset Search search bar "education" and then hit enter. See, now on the left side the dataset search lists sources where it found "education" data. On the right side it gives you the preview of the, all of the information that is associated with this particular database. So, in this case it found education statistics from World Bank open data and then you can explore this at Kaggle, and then it tells you when this dataset was updated, the dataset, who provided this dataset. There is a license associated usually, in this case it's unknown. What are the available formats to download, the description, contexts and acknowledgements.

And if we keep going we'll see that it gets data sets from different sources. So, for example this one from Morocco, we can click here it was last updated in December 7th, 2018 and then we can explore this at "catalogue.data.gov". And then we click and then we see the website that published the data and then we can read more about it and even see if there are links to access the database itself, or even talk to folks that are actually managing this dataset so it can contact that authority that is maybe managing this dataset. The same goes for example for this indication in India dataset at Kaggle, we can explore this at Kaggle, and then when you open the web page it'll show you all of the information that is associated with this, the license, how you can download it, the content and everything else that is associated with this.

So Google Dataset Search it's a new tool. It's relatively recent that we launched, so that is focused only on data sets there are published by different sources around the web.