

Also new this week: “Google Dataset Search”

Egon Willighagen 

Published September 8, 2018

Citation

Willighagen, E. (2018). Also new this week: “Google Dataset Search”. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/e05yz-mhy44>

Keywords

Data, Google

Abstract

There was a lot of Open Science news this week.

Copyright

Copyright © Egon Willighagen 2018. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

chem-bla-ics

There was a lot of Open Science news this week. The [announcement](#) of the [Google Dataset Search](#) was one of them:

Google Dataset Search Beta

Search for Data Sets



Try [boston education data](#) or [weather site:noaa.gov](#)

Of course, I first tried searching for “[RDF chemistry](#)” which shows some of my data sets (and a lot more):

The screenshot shows the Google Dataset Search interface. At the top, the search bar contains the query "RDF chemistry". Below the search bar, there are several search results. The first result is highlighted and shows a preview of the data source: "Data from: ChemPedia as RDF" from figshare.com, updated on Jan 11, 2016. Below this, there are three other results: "ChemPedia RDF" from datahub.io (updated Oct 10, 2013), "ChEMBL-RDF v13.5" from dataverse.nl (updated Aug 29, 2013), and "Canopy Chemistry (OTTER)" from datahub.io. On the right side of the page, there is a detailed view of the selected result, showing the same source information, a "Related Article" link, the figshare.com logo, and metadata: "Data set created Apr 14, 2013", "Data set updated Jan 11, 2016", and "Data set published Apr 14, 2013". It also lists the provider as "figshare" and the author as "Egon Willighagen".

It picks up data from many sources, such as [Figshare](#) in this image. That means it also works (well, sort of, as [Noel O'Boyle noticed](#)) for supplementary information from the [Journal of Cheminformatics](#).

It picks up metadata in several ways, among which [schemas.org](#). So, next week we'll see if we can get [eNanoMapper](#) extended to spit compatible JSON-LD for its data sets, called “bundles”.

Integrated with Google Scholar?

While the URL for the search engine does not suggest the service is more than a 20% project, we can hope it will stay around like Google Scholar has been. But I do hope they will further

chem-bla-ics

integrate it with Scholar. For example, in the above figure, it did pick up that I am the author of that data set (well, repurposed from an effort of [Rich Apodaca](#)), it did not figure out that I am also on Scholar.

So, these data sets do not show up in your Google Scholar profile yet, but they **must**. Time will tell where this data search engine is going. There are many interesting features, and given the amount of online attention, they won't stop development just yet, and I expect to discover more and better features in the next months. Give it a spin!