

# Facilitating Dataset Discovery in an Open Ecosystem

Measuring Development 2020 Conference

Yale Cong

Google Research





There are  
thousands of data  
repositories across  
the web...

**re3**data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES


 **DataCite**  
FIND, ACCESS, AND REUSE DATA

 [data.nodc.noaa.gov](https://data.nodc.noaa.gov)

 [figshare.com](https://figshare.com)

 [datadryad.org](https://datadryad.org)

 [catalog.data.gov](https://catalog.data.gov)

 Harvard Dataverse

 [zenodo.org](https://zenodo.org)


 [www.europeandataportal.eu](https://www.europeandataportal.eu)

 Kaggle

 [www.datamed.org](https://www.datamed.org)

 [data.world](https://data.world)

 [data.nasa.gov](https://data.nasa.gov)

 [data.opendatanetwork.com](https://data.opendatanetwork.com)

Google Research

# Introducing Dataset Search

A focused search engine  
that leverages dataset  
**metadata** from data  
providers

[q.co/datasetsearch](https://q.co/datasetsearch)

Google Dataset Search

# How do I include my datasets?

We rely on providers to use [schema.org](https://schema.org), an open, community-based standard for structured data.

The logo for schema.org, featuring the text "schema.org" in white lowercase letters on a dark red rectangular background.

schema.org

# Open Data Ecosystem

Data providers

**Publish**  
structured metadata  
using community  
standards

Data consumers

**Cite**  
data properly, like how  
we cite scientific  
publications

Developers

**Contribute**  
to expanding  
metadata for datasets

# Thank You!

Check out the tool at  
[g.co/datasetsearch](https://g.co/datasetsearch)

