Facilitating Dataset Discovery in an Open Ecosystem

Natasha Noy
Google, Inc.
Why Dataset Search?

Nature Scientific Data recommends 58 repositories

2,000 Data Repositories and Science Europe’s Framework for Discipline-specific Research Data Management

- data.nodc.noaa.gov
- catalog.data.gov
- data.world
- www.europeandataportal.eu
- data.nasa.gov
- www.datamed.org
- figshare.com
- zenodo.org
- data.opendatanetwork.com
- datadryad.org
What is Dataset Search?

It's a search engine over metadata
dataset name

description

temporal coverage

provider
What is Dataset Search?

It's a search engine over metadata from data providers.
Why schema.org?

- It's an open standard
- Adoption driven by use in real search products
- Embedded in HTML
- Anybody can read and crawl this metadata
  - And build tools over it
- It is really easy to add it. We promise!
Where is my dataset?

- Does the page have schema.org metadata?
- Has it been indexed?
- Do you have a sitemap?

Google Structured Data Testing Tool

sitemaps.org

Google Search Console
thousands of domains

millions of datasets

Google Dataset Search Beta

Search for Datasets
Next steps

Create a healthy data ecosystem

Data providers

- publish structured metadata using community standards

Data consumers

- cite data properly, much as we cite scientific publications

Developers

- contribute to expanding metadata for datasets