

# Finding Data

## Finding data: the Data State of Mind

- Question/hypothesis, then –
  - What would the data look like? – what variables would it contain, what would they be called?
  - Who would collect such data?
  - Where can I get hold of it?
  - If it doesn't exist, can I create it myself?

# 4 sources

Official portals

Unofficial portals

Google search

Scraping

## Finding data: portals

## Open Data Inception - 2600+ Open Data Portals Around the World

Try Education, France or Weather 

Clear a



opendatainception.io

## Finding data: government portals

data.gov.uk | Find open data

[Publish your data](#) [Documentation](#) [Support](#)

**BETA** This is a new service – your [feedback](#) will help us to improve it

# Find open data

Find data published by central government, local authorities and public bodies to help you build products and services



[Business and economy](#)

Small businesses, industry,

[Environment](#)

Weather, flooding, rivers, air

[Mapping](#)

Addresses, boundaries, land

## Finding data: government portals



[DATA](#) [TOPICS ▾](#) [IMPACT](#) [APPLICATIONS](#) [DEVELOPERS](#) [CONTACT](#)

# The home of the U.S. Government's open data

Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and [more](#).

## GET STARTED

SEARCH OVER [235,899 DATASETS](#)



[BROWSE TOPICS](#)

# Finding data: international portals

New to this site? [Start Here](#)

[DataBank](#) [Microdata](#) [Data Catalog](#) 

## World Bank Open Data

Free and open access to global development data

Search data e.g. GDP, population, Indonesia

Browse by [Country](#) or [Indicator](#)

### MOST RECENT

March 2020 global poverty update from the World Bank: New poverty estimates for 2018 [↗](#)

R. Andres Castaneda Aguilar, Tony Fujs, Christoph Lakner, Daniel Gerszon Mahler, Minh Cong Nguyen, Mar 31, 2020

### WHAT YOU CAN LEARN WITH OPEN DATA

Poverty headcount ratio at \$1.90 a day (2011 PPP) (% of population)

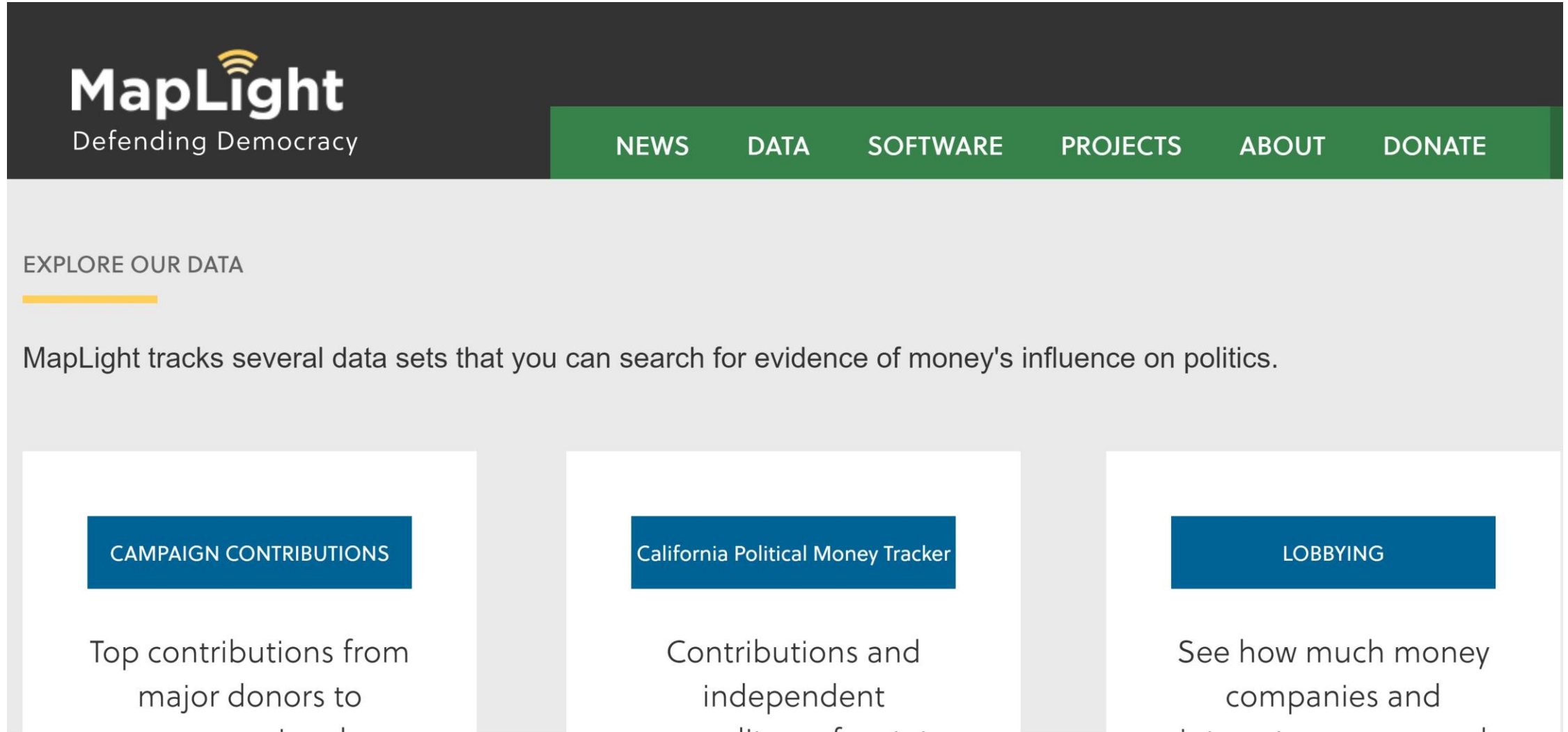


INTERNATIONAL  
DEBT STATISTICS

2020

 [Help / Feedback](#)

## Finding data: non-government portals



The screenshot shows the MapLight website interface. At the top, there is a dark grey header with the MapLight logo (a yellow signal icon above the text 'MapLight') and the tagline 'Defending Democracy'. To the right of the logo is a green navigation bar with white text links: NEWS, DATA, SOFTWARE, PROJECTS, ABOUT, and DONATE. Below the navigation bar, the main content area has a light grey background. It starts with the section 'EXPLORE OUR DATA' followed by a yellow horizontal line. A paragraph states: 'MapLight tracks several data sets that you can search for evidence of money's influence on politics.' Below this, there are three white rectangular boxes, each with a blue header and descriptive text. The first box is titled 'CAMPAIGN CONTRIBUTIONS' and describes 'Top contributions from major donors to'. The second box is titled 'California Political Money Tracker' and describes 'Contributions and independent'. The third box is titled 'LOBBYING' and describes 'See how much money companies and'.

**MapLight**  
Defending Democracy

NEWS DATA SOFTWARE PROJECTS ABOUT DONATE

EXPLORE OUR DATA

MapLight tracks several data sets that you can search for evidence of money's influence on politics.

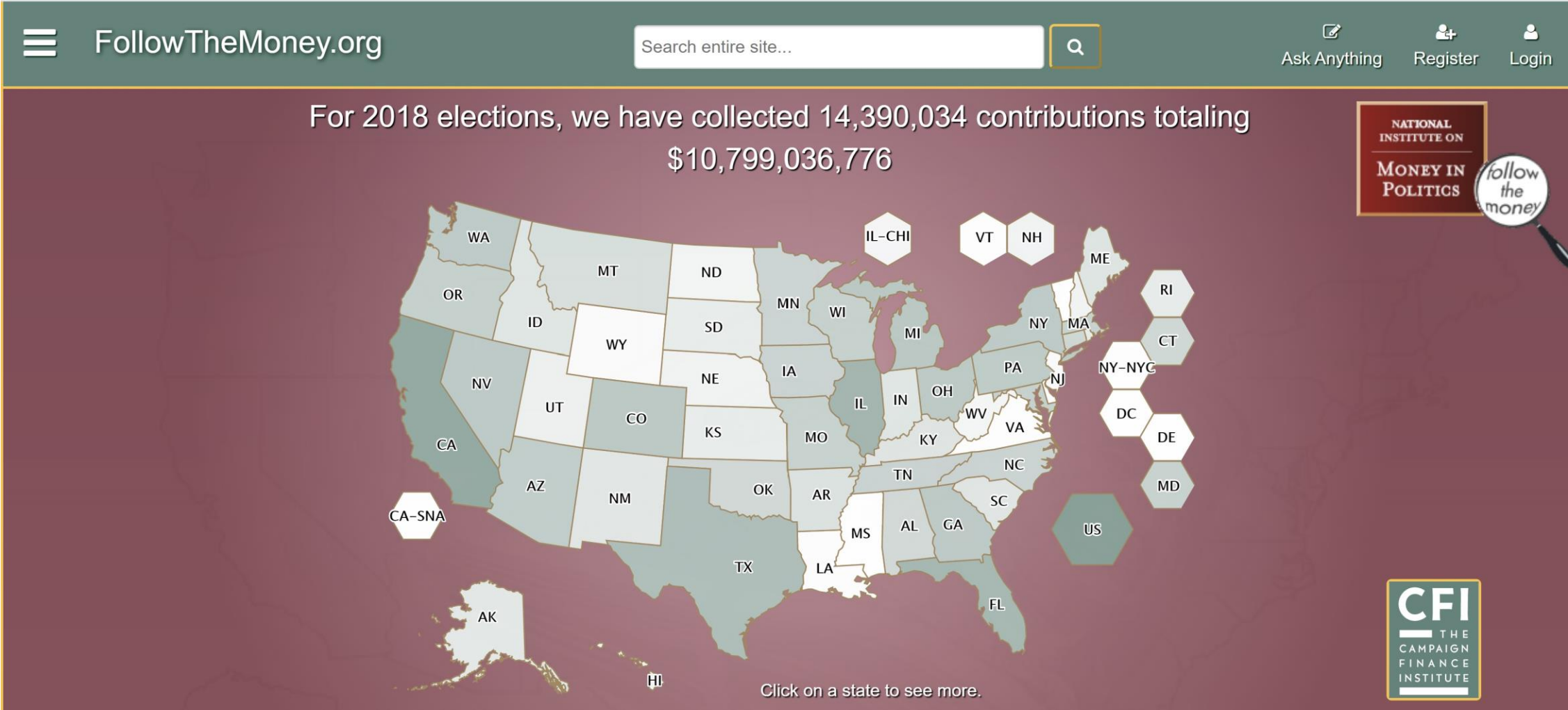
**CAMPAIGN CONTRIBUTIONS**  
Top contributions from major donors to

**California Political Money Tracker**  
Contributions and independent

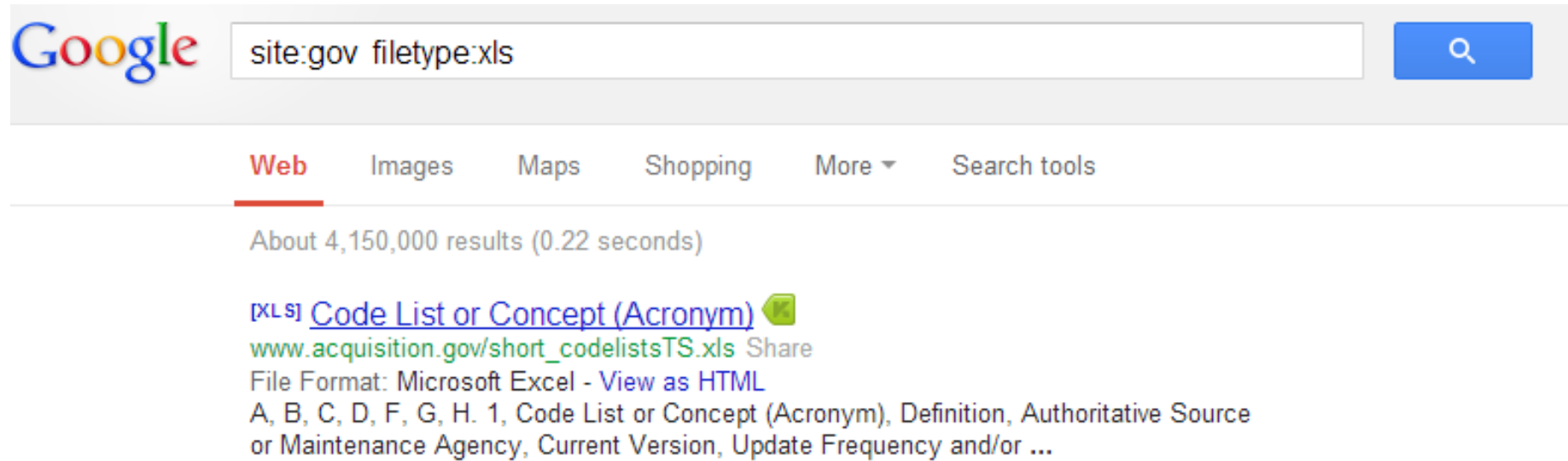
**LOBBYING**  
See how much money companies and



# Finding data: non-government portals




## Finding data: Google advanced

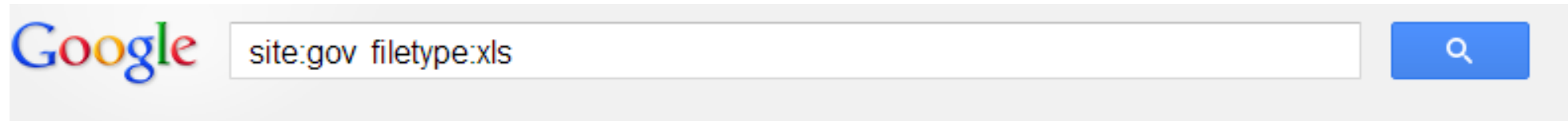


The screenshot shows a Google search interface. The search bar contains the text "site:gov filetype:xls". Below the search bar, the "Web" tab is selected. The search results indicate "About 4,150,000 results (0.22 seconds)". The first result is titled "[XLS] Code List or Concept (Acronym)" with a green Excel icon. The URL is "www.acquisition.gov/short\_codelistsTS.xls" and it includes a "Share" link. Below the URL, it says "File Format: Microsoft Excel - View as HTML". The description of the file is "A, B, C, D, F, G, H. 1, Code List or Concept (Acronym), Definition, Authoritative Source or Maintenance Agency, Current Version, Update Frequency and/or ...".

site:gov.uk ....ac.uk, edu, gov, br, fr, de, ba....

[XLS] TSB Monthly Cash Flow Projection   
www.dia.iowa.gov/tsbmonthlycashflow.xls

## Finding data: Google advanced



Web Images Maps Shopping More ▾ Search tools

About 4,150,000 results (0.22 seconds)

site:gov.uk ....ac.uk, edu, gov, br, fr, de, ba....

filetype:xls....csv, (also doc, ppt, pdf)

[Approps - Foreign Assistance.gov](#)  
[www.foreignassistance.gov/Full\\_ForeignAssistanceData.xls](http://www.foreignassistance.gov/Full_ForeignAssistanceData.xls)

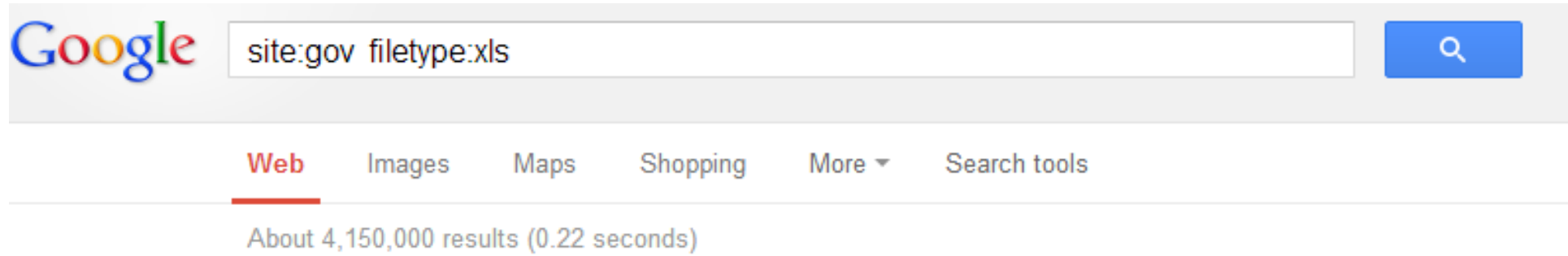
File Format: Microsoft Excel

A, B, C, D, E, F, G, H. 1, Planning Data. 2. 3. 4, Fiscal Year, Fiscal Year Type,  
Account Name, Agency Name, Operating Unit, Category, Sector, Amount ...

[XLS] [TSB Monthly Cash Flow Projection](#) 

[www.dia.icma.gov/tchcalncashflow.xls](http://www.dia.icma.gov/tchcalncashflow.xls)

## Finding data: Google advanced



site:gov.uk ....ac.uk, edu, gov, br, fr, de, ba....

filetype:xls....csv, (also doc, ppt, pdf)

inurl:....eg public, upload, 2015, inspection, report, dashboard, data....

## Finding data: Newsletters

No images? [Click here](#)



ISSUE 279 • March 31, 2020

Insight

### Why It's So Hard To Make A Good COVID-19 Model

31 March 2020 – #364: quantum of sollazzo

The data newsletter by @puntofisso

[Read it in your browser](#)

Made with 

# FAIR WARNING

**Fair Warning** - AirBnB, caucus cats,  
and the billionaire's billionaire

By [Sophie Warnes](#) • Issue #106 • [View online](#)

## Data Is Plural

by [Jeremy Singer-Vine](#)

A weekly newsletter of useful/curious datasets. Archive  
available as a spreadsheet [here](#). RSS feed [here](#).