



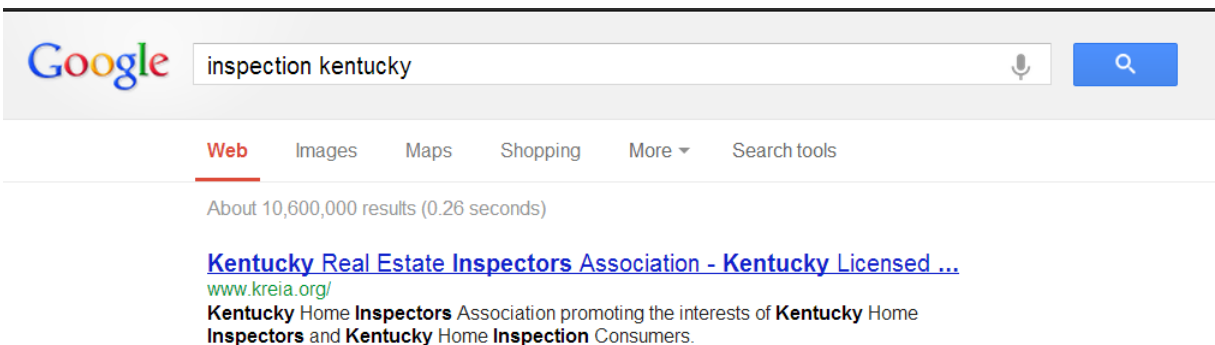
Finding data online

Jaimi Dowdell, IRE | Jaimi@ire.org

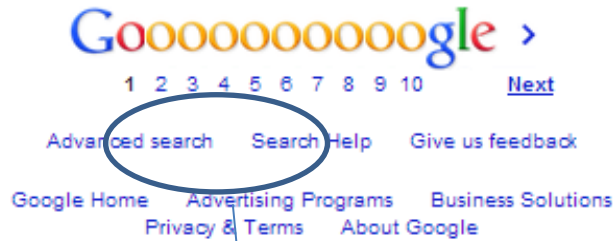
Data and data-based reporting are hot right now and the reason is quite simple: Incorporating these skills and this information into your work can produce better journalism. “I don’t have any data to work with” can’t be an excuse for not doing these types of stories because data are everywhere! In this tipsheet you’ll see some of my favorite tips, tricks and sites for finding data. You’ll also learn how you can easily get data into a spreadsheet once you’ve found it. Feel free to contact me with any questions: jaimi@ire.org.

Use Advanced Searches in Google

We’re used to just typing in whatever we want to find in the Google search bar. That works just fine when we’re trying to find lunch, but it doesn’t work as well when we’re trying to hunt down data.



Simple use of quotation marks can quickly narrow down your results. In addition, make sure you look at the advanced search in Google to see your options. It will help you save time and get better results. See the figure below:



Google

Advanced Search

Find pages with...

all these words:

this exact word or phrase:

any of these words:

none of these words:

numbers ranging from: to

To do this in the search box

Type the important words: telusiole vat teveler

Put exact words in quotes: "vat teveler"

Type OR between all the words you want: minilature OR standard

Put a minus sign just before words you don't want: -redneck -"Dale Gribble"

Put 2 periods between the numbers and add a unit of measure: 10..39 lb, \$200..\$500, 2010..2012

Then narrow your results by...

language: Find pages in the language you select.

region: Find pages published in a particular region.

last update: Find pages updated within the time you specify.

site or domain: Search one site (like wikipedia.org) or limit your results to a domain like .edu, .org or .gov

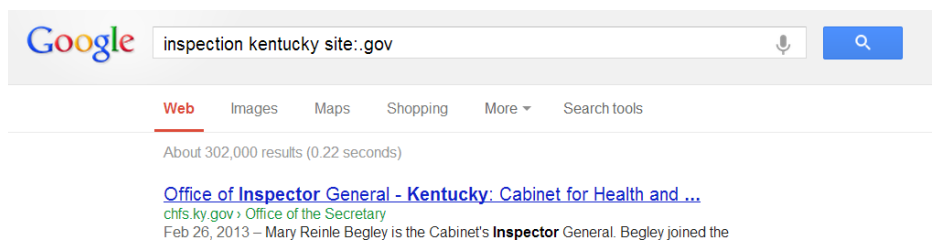
terms appearing: Search for terms in the whole page, page title, or web address, or links to the page you're looking for.

SafeSearch: Tell SafeSearch whether to filter sexually explicit content.

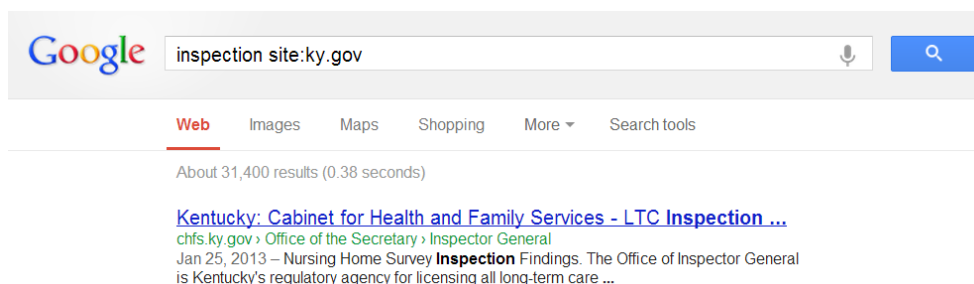
reading level: Find pages at one reading level or just view the level info.

file type: Find pages in the format you prefer.

You can change the original search to show only results from government domains



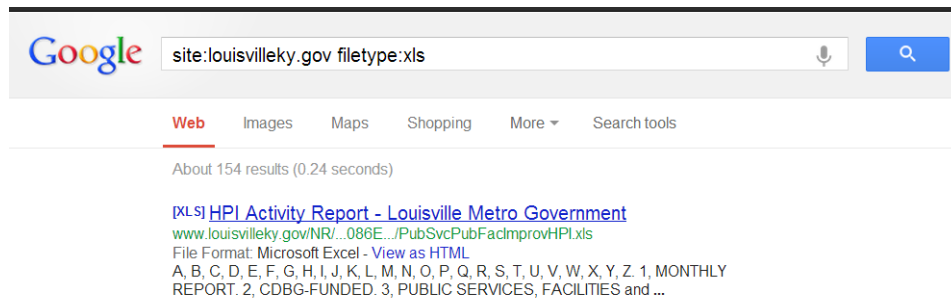
You can be even more specific. For example, instead of showing results for all government sites, now I am only looking for the word "inspection" in results from websites that end in "ky.gov"



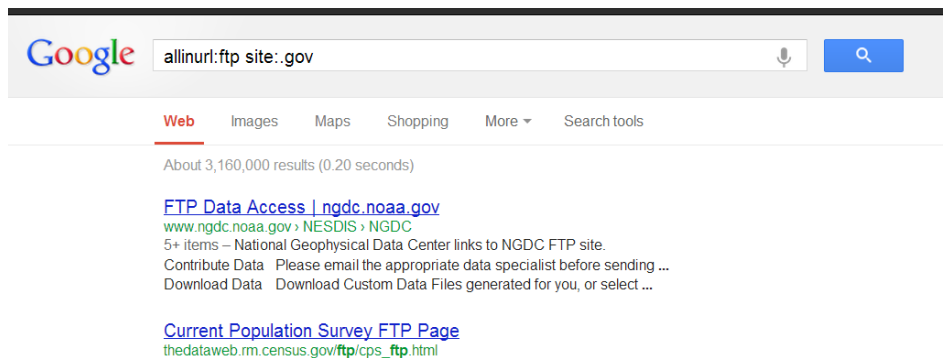
This is the same search as above, but I'm only looking for results that are spreadsheet files. You can use the "filetype:" narrower to specify results by filetype (spreadsheets, word docs, PDF files, PowerPoint presentations, etc.):



If you get rid of any search terms, you can find all of the spreadsheets (or any other file type) that Google has found on one specific site like this:



In addition, I like searching for ftp and other fun words in a URL. In the example below, I combine an "allinurl:" search with a domain narrower using "site:":



Clues that data exist

- Major Information Systems – request a list from government agency: example US Marshall’s list [here](#).
- Something in a “chart” format may mean that a database exists: example Texas offenders on death row [here](#).
- Online forms to submit information: example Texas Medical Board complaint form [here](#).
- Actually read a web site – don’t let it force you into predefined roles: example North Carolina Department of Revenue [here](#). (Read through the site. Look at the site map. Use their own search to search for keywords related to data: download, database, information system, submit, searchable, inspection, enforcement, Excel, etc.)

Retention schedules

- A plan to keep, maintain and dispose of records.
- Treat the retention schedule like a menu of documents
 - List from the [Council of State Archivists](#)
 - [Example](#) from Illinois Department of Children & Family Services

Things to keep in mind

- Never hesitate to request information you find in a more up-to-date format.
- Some searchable databases do not include the entire dataset.
- Just like everything else, you must report out this information.

Downloading and converting data

(Once you’ve found it, how do you get it somewhere you can use it?)

- In Excel format – example IRS Exempt Organizations [here](#).
- From the web – example Texas offenders on death row [here](#). (See tutorial beginning on page 8 of this handout)
- In text format – example from data.gov search on “inspection” [here](#).
- Use [DTA](#) to quickly download records – example Oklahoma [audit reports](#).
- PDF to Excel – try [cometdocs.com](#) (Note: IRE members can get a free premium Cometdocs account. Details [here](#))
- Use Web scraper for Chrome. Example: World Bank [Ineligible listings](#).
- [Creative searching](#) in forms. (Remember, for creative searching in search forms, try not using any search terms or use wild cards like * or %.)

Searchable and downloadable data that already exist online

- [OSHA](#) workplace safety inspections
- DOL [enforcement data](#)
- Federal contracts ([FPDS](#)) (ex. Border protection)
- [Recalls](#) (look at tabs for topics)
- [Economy at a Glance](#) from the Bureau of Labor Statistics
- Local Area Unemployment Statistics from the [BLS](#)
- [EPA](#) environmental data searches
- Fatality Analysis Reporting System data ([FARS](#))
- National sex offender registry ([Department of Justice](#))
- Transportation data from [BTS](#)
- Aircraft data (service difficulty reports, on-time data, aircraft registry, etc.)
 - NTSB [Aviation Accidents](#) (and more)
 - FAA [accidents/incidents](#) and service difficulty reports ([SDRs](#))
 - Flight tracker from [flightaware](#)
 - [On-time](#) statistics
 - Aviation [consumer reports](#) (PDF file)
 - [Bureau of transportation](#) statistics
- Nursing home, hospital and home health data from [Medicare](#)
- Firearms data from the [ATF](#)
- Migration data from the [IRS](#)
- Tax data from [IRS](#) (tax exempt [orgs](#))
- Mine Safety and Health Administration ([MSHA](#))
- National Agricultural [Statistics](#) Service
- The Census
 - [American FactFinder](#) for quick data profiles.
 - [Datasets](#) from IRE– download profiles, comparisons and more.
 - Get information on the Census from ASU’s Walter Cronkite [School of Journalism](#) and [IRE](#)
- Federal information
 - [Fedstats](#) – A to Z list of statistical data from the government.
 - [Data.gov](#) – Searchable catalogs of government data. (Search for business, contracts, inspections, etc.)

- USA Spending.gov or FedSpending.org – Federal spending data including grants, loans, contracts, etc. (Ex. Coca Cola)
- State and local data
 - State Government [Databases wiki](#) from the American Library Association
 - [BRB Publications](#) links to public records sites
 - [Portico](#)

Search for International information using US data

Federal Contracts

Search the FPDS site for contractors based in a country:

https://www.fpds.gov/fpdsng_cms/

You can narrow your search by any field in the federal contracts data. Data is exportable to CSV.

Energy administration information

International energy statistics:

<http://www.eia.gov/cfapps/ipdbproject/IEDIndex3.cfm?tid=5&pid=54&aid=4>

Net imports of petroleum to US:

http://www.eia.gov/dnav/pet/pet_move_neti_dc_nus-nda_mbbldpd_a.htm

Department of State

Weapons authorizations by country:

http://www.pmdtdc.state.gov/reports/655_intro.html

Federal Aviation Administration

Aircraft registry: http://registry.faa.gov/aircraftinquiry/country_inquiry.aspx

Securities and Exchange Commission

EDGAR search by country:

<http://sec.gov/edgar/searchedgar/webusers.htm>

Internal Revenue Service

IRS tax exempt organizations – international file can be found here in Excel:

<http://www.irs.gov/taxstats/charitablestats/article/0,,id=97186,00.html>

Legistorm

For foreign gifts and trips by U.S. lawmakers: www.legistorm.com

Foreign lobbyists

Search by country: <http://www.fara.gov/quick-search.html>

Department of Homeland Security

Removals and returns:

<http://www.dhs.gov/files/statistics/publications/YrBk10En.shtm>

Sources for international data

- [UNdata](http://data.un.org) - data.un.org – This is a portal to a collection of international databases. The data is searchable and downloadable. This site is good at pointing out where the data came from so you can track it down at the source.
- [International Data Base from the U.S. Census Bureau](#) - Demographic data including growth rate, population and gender. This site also has a ranking option and describes, in detail, the methodology behind the data.
- [International trade and tariff data from the World Trade Organization](#) - Statistics on international trade as well as country profiles.
- [The UJIMA Project](#) - This site brings together databases and documents in the hopes to bring more transparency to African countries. Data includes information on spending by the U.S. Agency for International Development, foreign government lobbying of U.S. politics, weapon sales, disbursements from the Global Fund to fight HIV/Aids , and much more. This site also links to the source of the data so you can see or download the original information.
- **Data from the World Bank** - Search for data by topic or country. Topics include environment, energy & mining, health, economic policy, and more. Use the “[Data Catalog](#)” to download data sets.
- [NationMaster](#) - Much like Wikipedia, this site is a good way to get quick information and find sources for data. It contains searchable encyclopedic and statistical information for countries. It also has a lot of information about sports. Be wary, though, and make sure you always check the source of the information – sometimes it is Wikipedia - nonetheless, a good place to start.
- [World Health Organization Global Health Observatory \(GHO\) database](#) - Find health statistics around the world for things such as mortality, life expectancy, poverty and more. Individual country statistics are also available and downloadable.

Importing into Excel: HTML charts

Often when you search for information online or peruse a site, you come across a chart such as this:

2011 WNV Activity by County

(2012-03-20) - Source: <http://www.westnile.ca.gov>

Total Positive Counties: 36

County	Human Cases	Horses	Dead Birds	Mosquito Samples	Sentinel Chickens	Squirrels
State Totals	158	15	688	2087	391	24
Alameda	-	-	-	-	-	1
Alpine	-	-	-	-	-	-
Amador	1	-	-	-	-	-
Butte	3	-	-	1	20	-
Calaveras	-	-	-	-	-	-
Colusa	-	-	-	-	-	-
Contra Costa	3	-	38	7	-	-
Del Norte	-	-	-	-	-	-
El Dorado	1	-	2	-	-	-
Fresno	9	5	15	123	-	-
Glenn	1	-	2	-	5	-
Humboldt	-	-	-	-	-	-
Imperial	-	-	-	3	3	-
Inyo	-	-	-	2	-	-
Kern	18	1	4	389	140	-
Kings	1	-	9	63	-	-
Lake	-	-	1	3	-	-
Lassen	-	-	-	-	-	-
Los Angeles	58	1	227	467	86	15
Madera	2	-	3	5	-	-
Marin	-	-	-	-	-	-
Mariposa	-	-	-	-	-	-
Mendocino	-	-	-	-	-	-
Merced	1	3	9	11	15	-
Modoc	-	-	-	-	-	-
Mono	-	-	-	-	-	-
Monterey	-	-	-	-	-	-
Napa	-	-	-	-	-	-
Nevada	-	-	1	-	-	-
Orange	10	-	49	91	-	1

Anytime you see data in nice, neat columns and rows you should think several things:

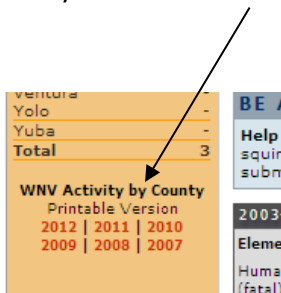
- 1) You can get this data into Excel fairly quickly and simply.
- 2) This is summary data aggregated by county; you can always request the data they used to create this summary chart.
- 3) The government agency may have much more information than shown.
- 4) The government agency may have more current information than shown.

The newer versions of Excel have made it incredibly easy to grab data from the Web. Try it out yourself by following the steps below:

Note: The following instructions will work only for Windows versions of Excel.

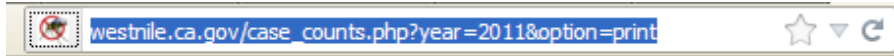
Step 1) Identify a Web page you'd like to pull into Excel. For example, go to California's West Nile Virus page here: <http://westnile.ca.gov/>.

Once you're at the page, scroll down and look at the left-hand navigation until you see "WNV Activity by County" as shown here:



Select the 2011 printable version and you'll be taken to a chart as shown in the beginning of this document. This chart definitely looks like a spreadsheet with columns and rows.

Step 2) Highlight the URL for the page and copy it:



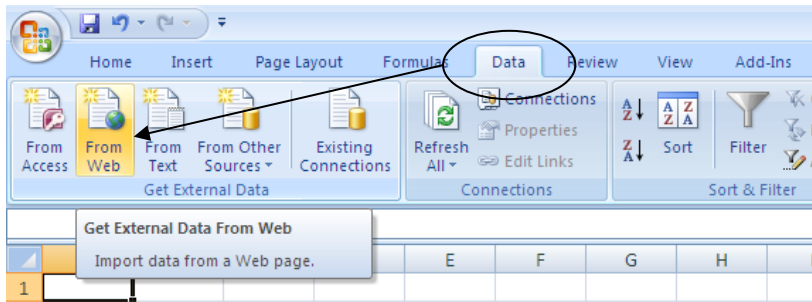
2011 WNV Activity by County

(2012-03-20) - Source: <http://www.westnile.ca.gov>

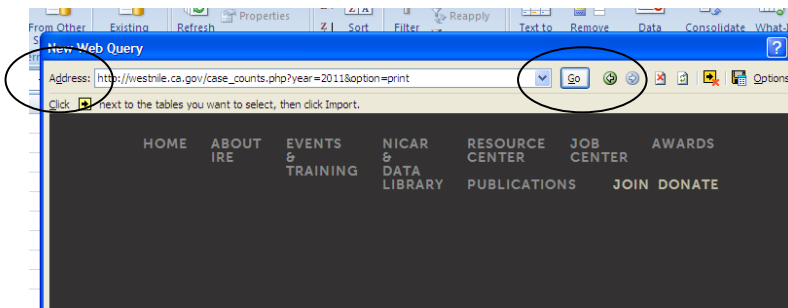
Total Positive Counties: 36

Step 3) Open up a blank worksheet in Excel.

Step 4) Click on the "Data" tab. Under "Get External Data" look for "From Web." Click on that icon.



Step 5) When the "New Web Query" window opens, paste the URL in the "Address" box and click on "Go."



Step 6) That should bring up the page. Notice the yellow arrows throughout the page. These arrows indicate different html tables in the page. If you want to pull in the entire page, simply click "Import." If you'd like to select certain tables then click on the arrow next to the table(s) you'd like and the arrows will become green checkmarks. For this example, just click "Import."

Address: http://westnile.ca.gov/case_counts.php?year=2011&option=print

Click next to the tables you want to select, then click Import.

2011 WNV Activity by County

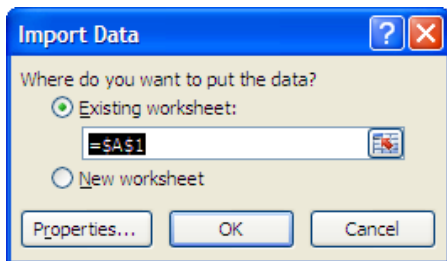
(2012-03-20) - Source: <http://www.westnile.ca.gov>

Total Positive Counties: 36

County	Human Cases	Horses	Dead Birds	Mosquito Samples	Sentinel Chickens	Squirrels
State Totals	158	15	688	2087	391	24
Alameda	-	-	-	-	-	1
Alpine	-	-	-	-	-	-
Amador	1	-	-	-	-	-
Butte	3	-	-	1	20	-
Calaveras	-	-	-	-	-	-
Colusa	-	-	-	-	-	-
Contra Costa	3	-	28	7	-	-
Del Norte	-	-	-	-	-	-
El Dorado	1	-	2	-	-	-
Fresno	9	5	15	123	-	-
Glenn	1	-	2	-	5	-
Humboldt	-	-	-	-	-	-
Imperial	-	-	-	3	3	-
Inyo	-	-	-	2	-	-
Kern	18	1	4	389	140	-
Kings	1	-	9	63	-	-
Lake	-	-	1	3	-	-
Lassen	-	-	-	-	-	-

Buttons: Import, Cancel

Step 7) Excel will ask you where you'd like to put the data. This is why we started with a new worksheet. Just click "OK" here and your data should appear in the worksheet.



Step 8) Save the sheet as an Excel file. I also like to add a worksheet to the workbook that I name "Notes." In this sheet I paste the download link for the data and I list the download date for future reference.

More Practice: Pick and choose

For more practice, try importing data from the Minnesota Department of Natural Resources. Go to their enforcement division page here: <http://www.dnr.state.mn.us/enforcement/index.html>.

Select "Game & fish license revocations" or just go here:

<http://www.dnr.state.mn.us/enforcement/revocations.html>

You should see a page similar to this:

www.dnr.state.mn.us/enforcement/revocations.htm

Minnesota Department of Natural Resources

Recreation | Destinations | Nature | Education / safety | Licenses / permits / regs.

Home > Enforcement >

License Revocations are governed by [MS. 97A.421](#).

Individuals listed have been convicted of violating Minnesota Game & Fish laws.

During their revocation period these individuals are also prohibited from purchasing licenses in member states in accordance with a Wildlife Violator Compact with those states. The member states include: Alabama, Alaska, Arizona, California, Colorado, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maryland, Michigan, Minnesota, Mississippi, Missouri, Montana, Nevada, New Mexico, New York, North Dakota, Ohio, Oregon, South Dakota, Tennessee, Utah, Washington, West Virginia, Wisconsin, and Wyoming.

Big Game **Small Game** **Fishing** **Trapping**

Big Game

Last Name	First Name	Middle Name	City	State	Birth Date	Conviction Date	Revocation Ends
AAMODT	CLAYTON	DENN	WANIMINGO	MN	10-Aug-1979	21-May-2009	21-May-2012
AGUILAR	ROLANDO	GARCIA	BLOOMINGTON	MN	25-Sep-1972	09-Dec-2011	09-Dec-2014
ALLEN	TANYA	ANN	CASS LAKE	MN	13-Jul-1977	02-May-2011	02-May-2012
ANDERSON	JEREMY	CLINTON	FARIBAULT	MN	09-May-1982	05-Oct-2011	05-Oct-2014
ANDERSON	STEPHEN	COLE	ROSEAU	MN	27-Jul-1988	11-May-2009	11-May-2012
ANGERMAN	JASON	PATRICK	HASTINGS	MN	12-Mar-1979	04-Aug-2009	04-Aug-2012
BAILEY	KIRK	DOUGLAS	FLOODWOOD	MN	26-Jan-1968	28-Jun-2010	28-Jun-2013
BANITT	JAMIE	ALLEN	ZUMBROTA	MN	29-Nov-1975	24-Feb-2010	24-Feb-2013
BARTH	WILLIAM	DEAN	BOVEY	MN	08-Nov-1971	01-Feb-2010	01-Feb-2013

Notice that there are different sections, or html tables for “Big Game,” “Small Game,” “Fishing” and “Trapping.” Go through the import steps listed above and notice that there will be yellow arrows at each of these sections in the New Web Query window. Play around with importing all or portions of this data to get a better understanding of how the import functions.