



OPEN
KNOWLEDGE
FOUNDATION
DEUTSCHLAND

THE STORY HUNT: UNCOVER THE EU

Getting & Finding Data



OpenBudgets.eu



DATENSCHULE

Getting & Finding Data

Civil society campaigns or journalistic research usually begin with a hypothesis or a particular question. Yet, in order to test the hypothesis it often lacks the sources and information. The right dataset can thus be the crucial asset to prove a story right. Finding & getting good sources is therefore an important skill in data driven journalism. This learning material is a gives an overview on good practices and and useful tools on how to find & get data from the web.

Table of Content

1. Data Portals
2. Freedom of Information
3. Using Search Engines strategically
4. Web - API 's
5. Scraping the Web

Data Portals

Many government institutions provide data via open data portals or repositories. More frequently they become updated on a regular basis by government officials. Below find some portals important for European Affairs.

List of Data - Portals:

The biggest data portal for the European Union is:

- The European Data Portal: <https://europeandataportal.eu/>

The biggest European data repository for budget data is:

- Open Spending: <https://openspending.org/>

Other interesting, freely available data could be:

- Eurostat: <http://ec.europa.eu/eurostat/de>
- The Cohesion Fund: <https://cohesiondata.ec.europa.eu/>
- European Farm Subsidies: <http://farmsubsidy.openspending.org/DE/>
- World Bank Data: <http://data.worldbank.org/>
- The Data Hub: <https://datahub.io/>
- Data from the UN: <http://data.un.org/>
- OECD Stats: <http://stats.oecd.org/Index.aspx>

Freedom of Information

Under the Freedom of Information Act (FOIA), every person has a right to request any recorded information held by a public authority, such as a government department or local council. Data can also be requested via the specific FOIA Platforms.

Helpful Resources:

- On the European Level you can send a FOIA request via:
 - AsktheEU: <https://www.asktheeu.org/>
- Although not fully complete, Alaveteli provides an overview of FOIA Platforms worldwide:
 - Alaveteli Deployments: <http://alaveteli.org/deployments/>
- For Germany, the Open Knowledge Foundation set up the Platform FragDenStaat:
 - FragDenStaat: <https://fragdenstaat.de/>

Using Search Engines strategically

Sometimes data portals are actually not that easy to find. Yet, the right usage of search engines can help to uncover rather hidden data sets.

A few useful tricks are:

- Take search terms as specific as possible
- Use operators to exclude (-) or include (+) certain terms
- Use quotation marks to obtain exact results for search terms
- By specifying the filetype (e.g. "filetype:XLS") results limited to datasets

Further Resources for Search Engine Tricks:

- TechRepublic: <http://www.techrepublic.com/blog/10-things/10-tips-for-smarter-more-efficient-internet-searching/>
- SearchEngineWatch: <https://searchenginewatch.com/>
- BigDataUniversity: <https://bigdatauniversity.com/courses/introduction-data-journalism/>
- The Data Journalism Handbook includes a chapter with many examples & useful tricks on getting datasets.

WEB - API's

An API (Application programming interface) is a specially designed accessing point in order to obtain data.

Twitter, Facebook and many more provide APIs from which a user can request data in a structured way from a large database. For many data-driven projects the obtained data can be the starting point for a great story.

Example:

- The APIs from the New York Times lets you research articles back to 1851 <https://developer.nytimes.com/>
- The Twitter API lets you create tweets, research profiles and information on the follower of an account: <https://dev.twitter.com/rest/public>
- The Organized Crime and Corruption Reporting Projects (OCCRP) set up the Investigative Dashboard containing many leaked documents and those used in previous research on crime and corruption.

Scraping the Web

Web scraping describes the automated process of downloading data. It is especially useful, to avoid tedious copy and pasting. It works well for obtaining data from a page which is structured in a consistent way. A newspaper repository is a good example of such. There are many tools and plugins which help you scraping the web.

Google Sheets:

In html, data is structured but not in an easily usable format. One common tool that can be used to get data from html pages is Google Sheets. Google Sheets is the online version of Excel that can be collaboratively used and edited. Through using the “=importHTML” function, tables and lists can be easily imported.

Google Chrome Scraper Plugin:

Another useful tool for obtaining data is the Google Chrome Scraper Plugin. The a browser extension is a point and click tool that allows you to easily select elements on a web page and then extract the information into a data set. This Youtube Video explains how to use the Google Chrome Scraper Plugin: <https://www.youtube.com/watch?v=oCp9lCdSpZI&t=155s>

About Us



The Open Knowledge Foundation Germany is a nonprofit organization that advocates open knowledge, open data, transparency, and civil participation.



OpenBudgets is an EU funded project, aiming at supporting journalists, civil society organisations, NGOs, citizens and public administrations, by providing an overview of public spending, as well as tools and appropriate data and stories to advocate and fight for fiscal transparency.



School of Data Germany helps non-profit organisations, civil rights defenders and activists to understand and use data and technology effectively to increase their impact on societal challenges.



OPEN
KNOWLEDGE
FOUNDATION
DEUTSCHLAND



THE STORY HUNT: UNCOVER THE EU