

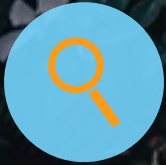
STORY HUNT

Workshop #2: Getting & Finding Data

Today's Program:

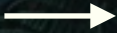
- The Data Pipeline
- Sources to get & find data
- Strategic Searching
- Google Sheets Basics

The Data Pipeline: Theory



Finding & Getting

- * Where to find Data?
- * Data portals
- * Freedom of Information Requests
- * Webs-scraping



Cleaning

- * Creating a structured Dataset



Analyzing

- * How do I spot:
- * Patterns & Trends ?
- * Outlier ?
- * Connections between Columns, Datasets
- * relation between Data & Context



Visualizing

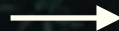
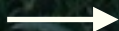
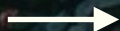
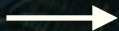
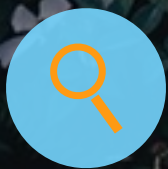
- * How to visualize the essence of a finding:
- * appealing
- * informative



Communicating

- * What's the right:
- * Format?
- * Medium?
- * Target group?

The Data Pipeline: Tools



Finding & Getting

- * Google Sheets
- * Webscraper
- * Unbubble
- * Tabula

Cleaning

- * Open Refine
- * Data Wrangler
- * Google Sheets / Excel
- * Python, R

Analyzing

- * Google Sheets / Excel
- * Python (Pandas)
- * R Studio

Visualizing

- * Infogr.am
- * Tableau Publique
- * Cato
- * ..Excel

Communicating

- * Timeline.js
- * Atavist

Sources to get & find data

EU - Data Portals

- The European Data Portal:
<https://europeandataportal.eu/>
- Eurostat:
<http://ec.europa.eu/eurostat/de>
- The Cohesion Fund:
<https://cohesiondata.ec.europa.eu/>
- European Farm Subsidies:
<http://farmsubsidy.openspending.org/DE/>

Other helpful sources:

- Lobbyfacts:
<https://lobbyfacts.eu/>
- RevolvingdoorWatch:
<https://corporateeurope.org/revolvingdoorwatch?>
- Organized Crime and Corruption Reporting Project (OCCRP):
<https://data.occrp.org/>
- OPEN CORPORATES:
<https://opencorporates.com/>
- OPEN OWNERSHIP:
<http://openownership.org/>

Do you know of other good sources?

Strategic Searching

Search operators

- “ ” = Only get results for the exact search term
- **OR / AND** = To get results in which terms are included
- - = Exclude specific search-results
- **filetype:** = Narrow search results to specific filetype (PDF, XLS, etc.)
- **site:** = Narrow search results to specific domain or e.g. country

GOOGLE SHEETS BASICS

- Google Sheets Formula:
<https://support.google.com/docs/table/25273?hl=en>
- RevolvingDoorWatch:
<https://corporateeurope.org/revolvingdoorwatch?>

NEXT WEEK: Scraping & Cleaning Data