Computational Analysis of Big Data

Week 3

Getting data

Overview

The data is already there

The data is already there

- Company data can give insight into your own **research** questions
- Market data can inform about **trends** that you may want to react upon
- You can **sell it**
- You can **combine multiple sources** and get unique insight that no one else has

Why do we want *other people's* data?

- Company data can give insight into your own **research** questions
- Market data can inform about **trends** that you may want to react upon
- You can **sell it**
- You can **combine multiple sources** and get unique insight that no one else has

How Facebook could stop a disease outbreak

January 3, 2018

Credit: National Cancer Institute

Facebook accounts and telephone records can be used to pinpoint the best individuals to vaccinate to stop a disease outbreak in its tracks, researchers said Wednesday.

Such people would be "central" in their social networks, and thus likelier to spread disease-causing germs from one group to another.

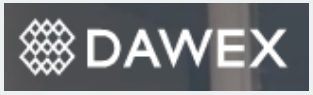
Assuming there is an outbreak, and not enough vaccines for every person in the world, immunising these well-connected individuals would remove social "bridges" by which germs can spread, experts wrote in the *Journal of the Royal Society Interface*.

Why do we want *other people's* data?

- Company data can give insight into your own **research** questions
- Market data can inform about **trends** that you may want to react upon
- You can **sell it**
- You can **combine multiple sources** and get unique insight that no one else has

Why do we want *other people's* data?

- Company data can give insight into your own **research** questions
- Market data can inform about **trends** that you may want to react upon
- You can **sell it**
- You can **combine multiple sources** and get unique insight that no one else has



Why do we want *other people's* data?

- Company data can give insight into your own **research** questions
- Market data can inform about **trends** that you may want to react upon
- You can **sell it**
- You can **combine multiple sources** and get unique insight that no one else has

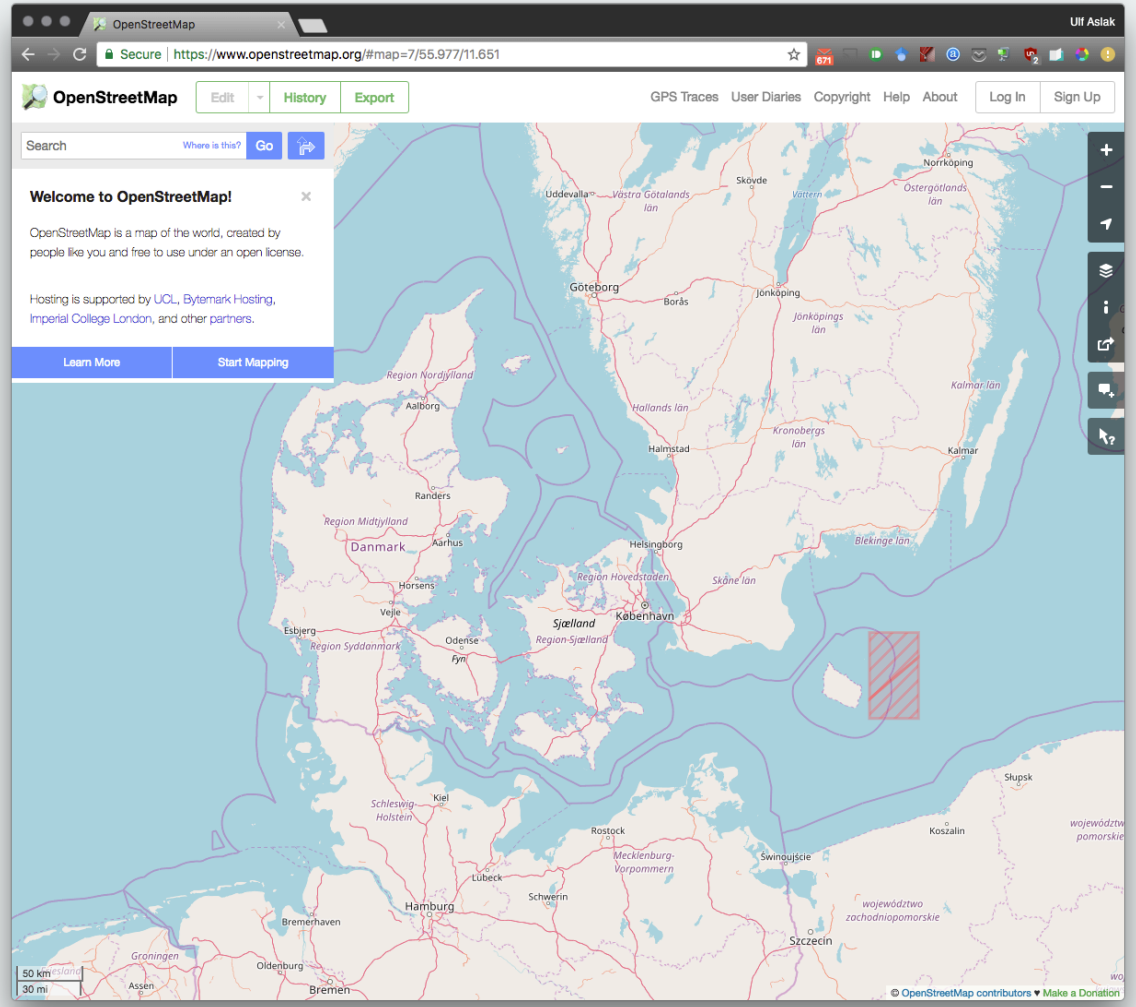
There are three main ways of getting existing data

- Get **open data** from public institutions, researchers or data sharing sites.
- Request it from someone's **API**. Is very easy, but usually has limits.
- Forcefully take it by **scraping** it from a website

Open data

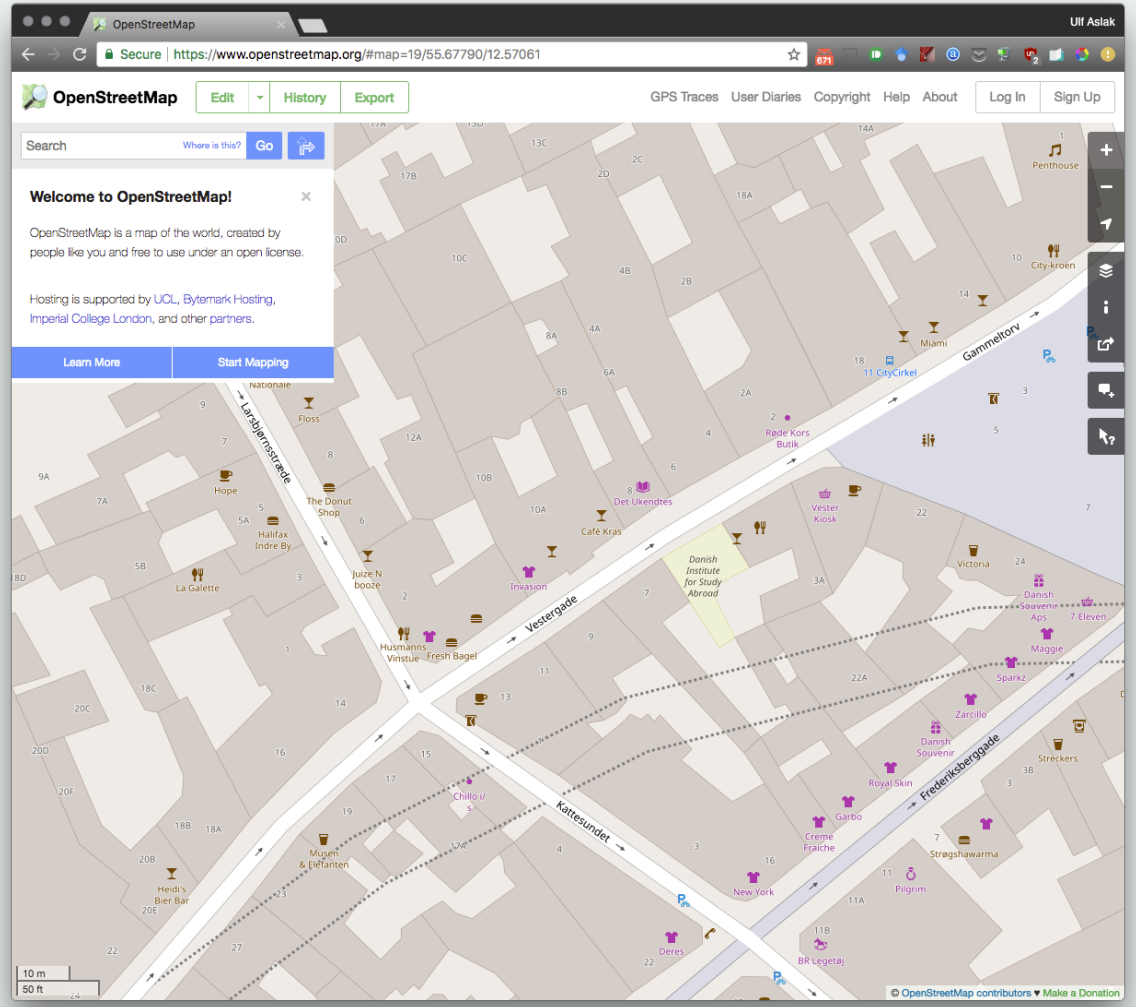
Open data

- **Geographical data**
- City data
- Political data
- Research data
- Competition datasets
- Transactional data



Open data

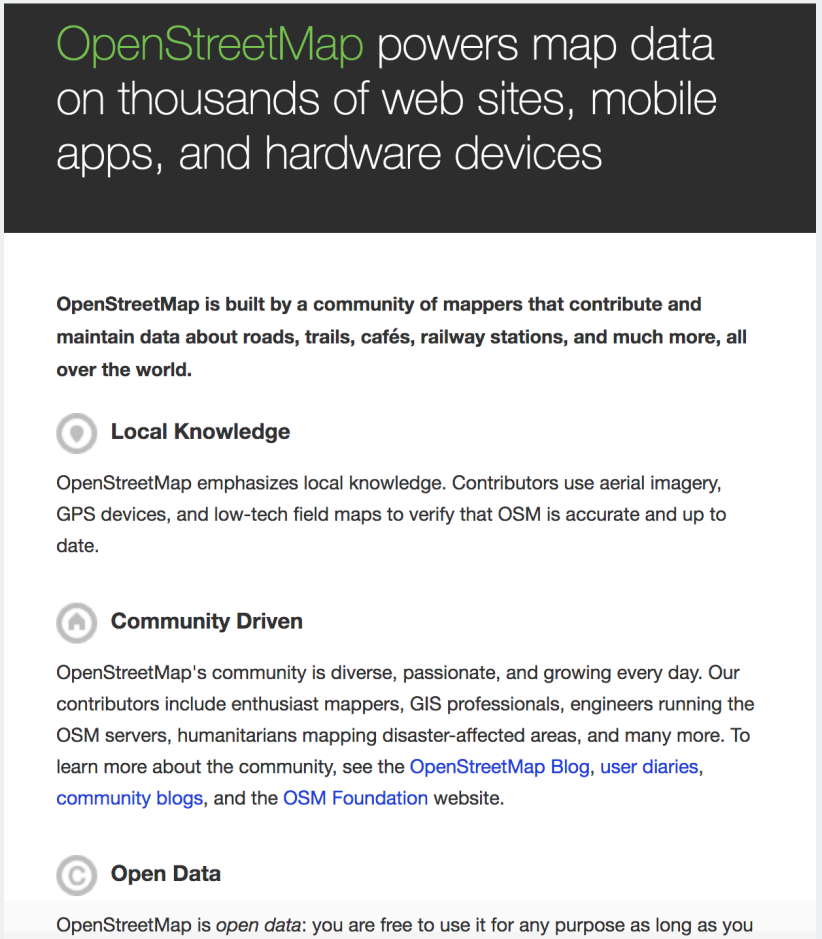
- **Geographical data**
- City data
- Political data
- Research data
- Competition datasets
- Transactional data



[Learn How to Map in OpenStreetMap]

Open data

- **Geographical data**
- City data
- Political data
- Research data
- Competition datasets
- Transactional data



OpenStreetMap powers map data on thousands of web sites, mobile apps, and hardware devices

OpenStreetMap is built by a community of mappers that contribute and maintain data about roads, trails, cafés, railway stations, and much more, all over the world.

Local Knowledge

OpenStreetMap emphasizes local knowledge. Contributors use aerial imagery, GPS devices, and low-tech field maps to verify that OSM is accurate and up to date.

Community Driven

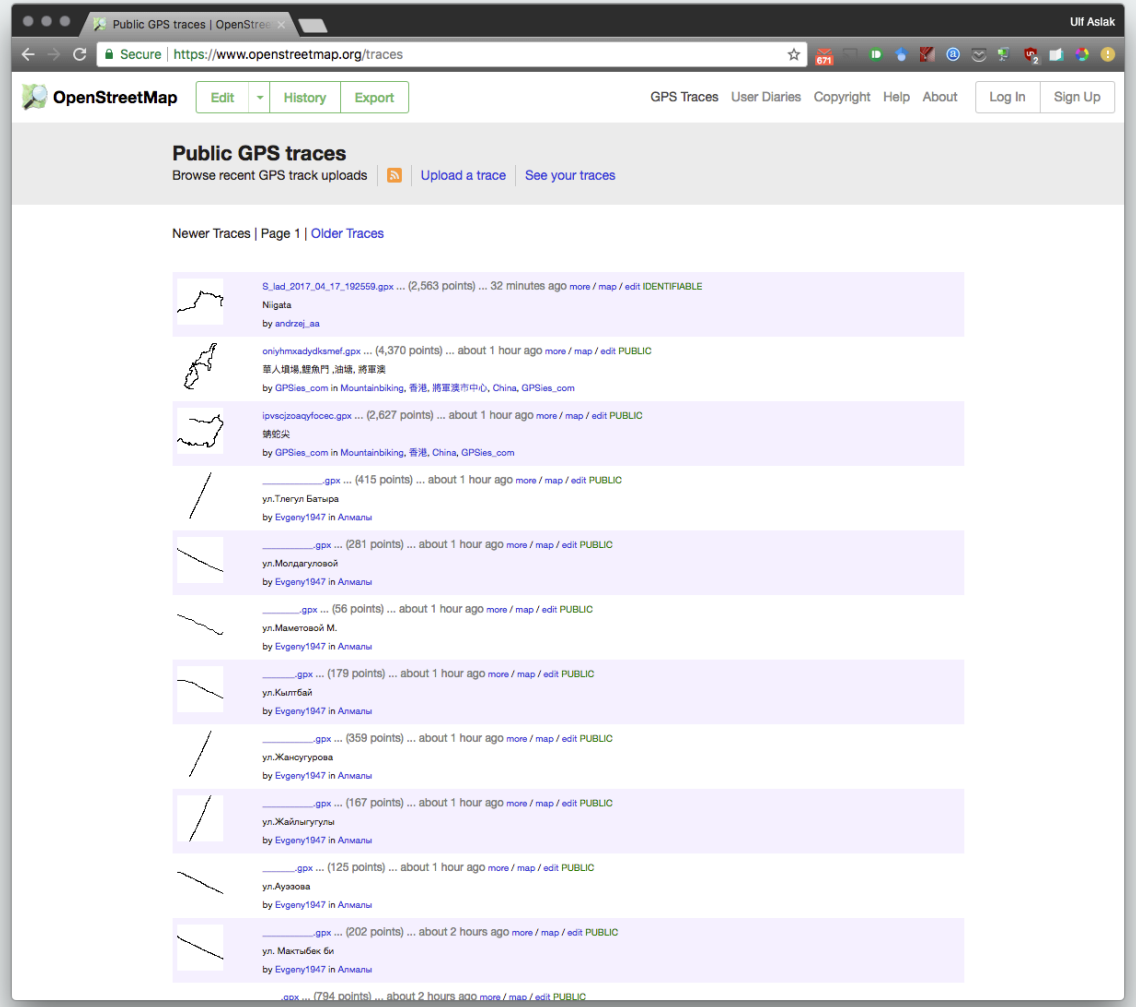
OpenStreetMap's community is diverse, passionate, and growing every day. Our contributors include enthusiast mappers, GIS professionals, engineers running the OSM servers, humanitarians mapping disaster-affected areas, and many more. To learn more about the community, see the OpenStreetMap Blog, user diaries, community blogs, and the OSM Foundation website.

Open Data

OpenStreetMap is *open data*: you are free to use it for any purpose as long as you

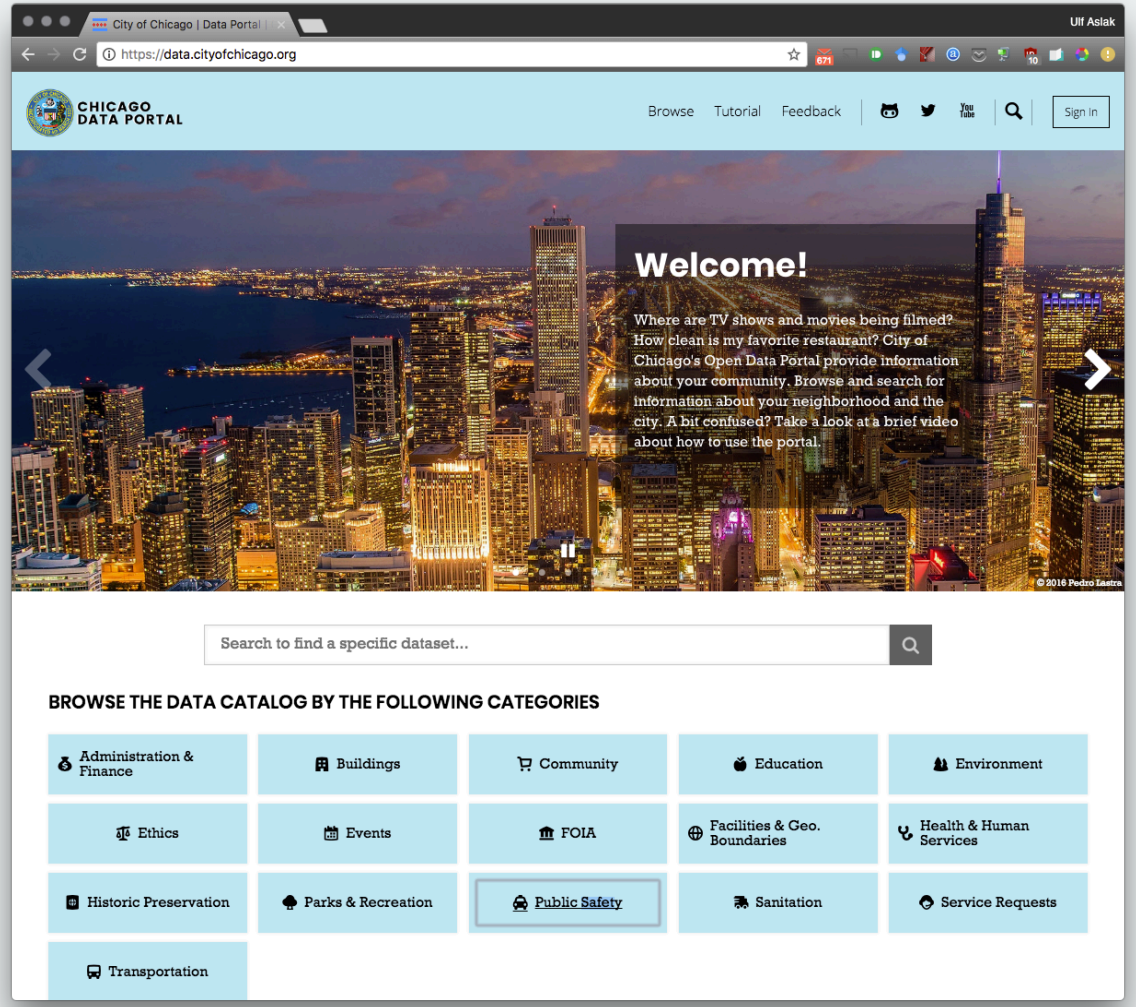
Open data

- **Geographical data**
- City data
- Political data
- Research data
- Competition datasets
- Transactional data



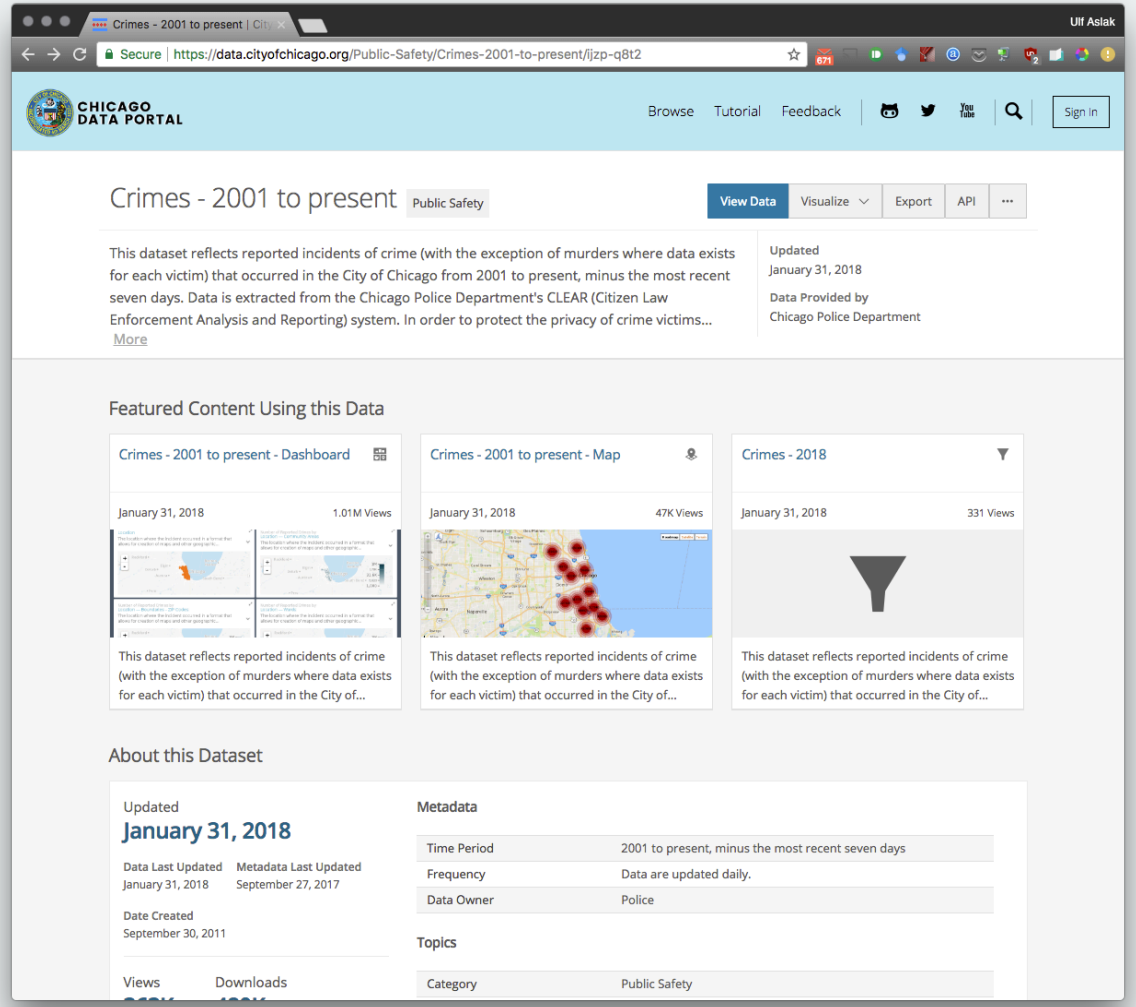
Open data

- Geographical data
- **City data**
- Political data
- Research data
- Competition datasets
- Transactional data



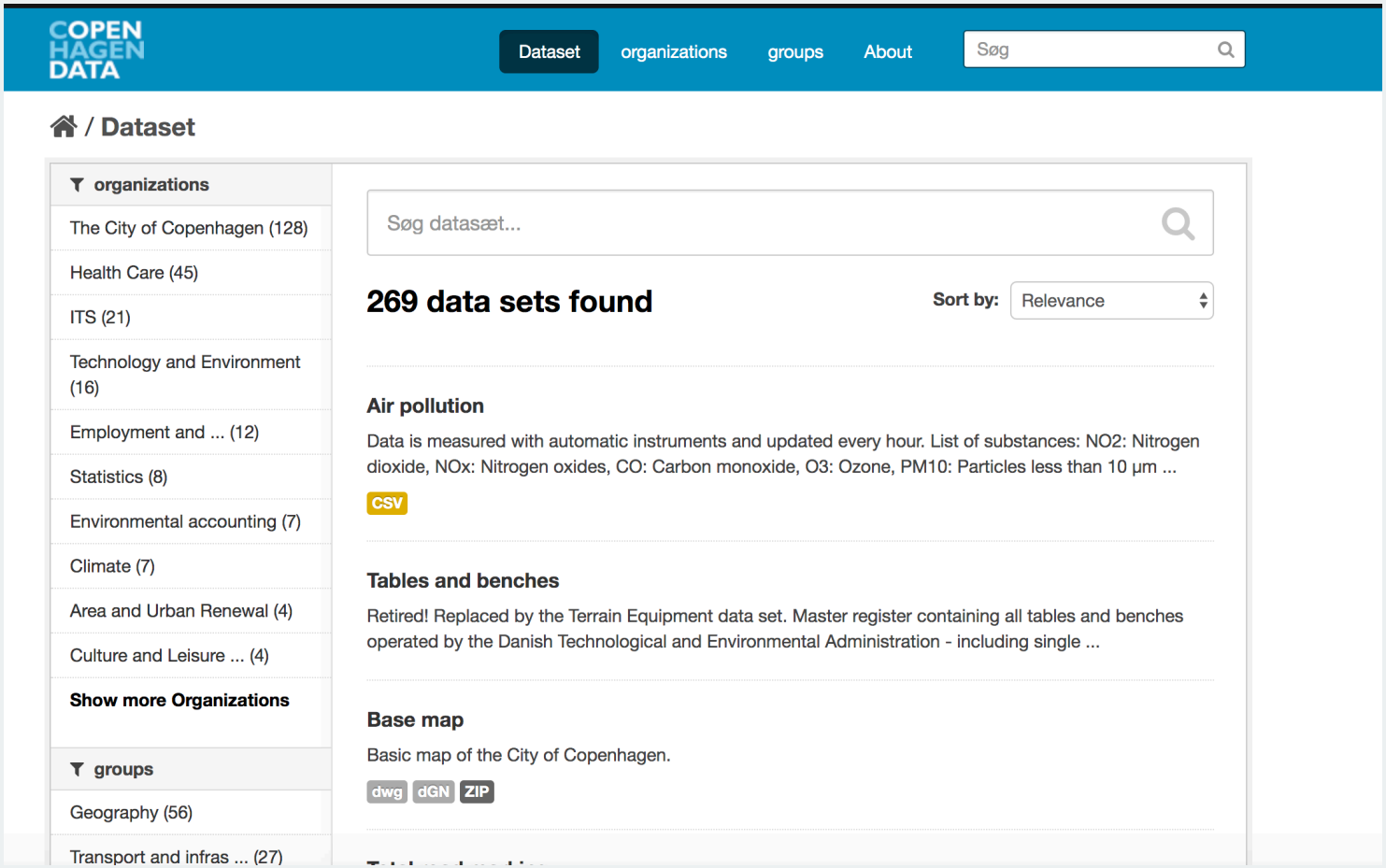
Open data

- Geographical data
- **City data**
- Political data
- Research data
- Competition datasets
- Transactional data



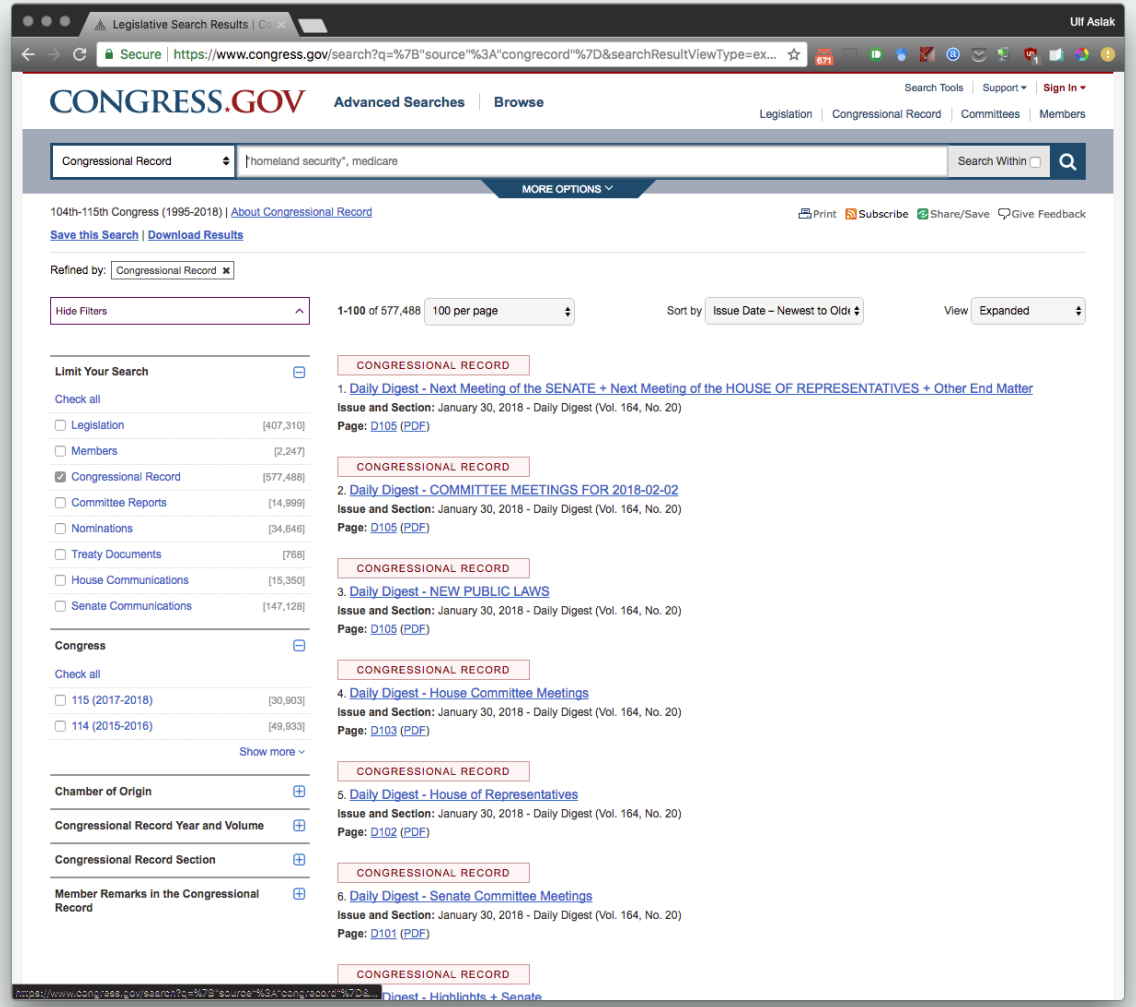
Open data

- Geographical data
- **City data**
- Political data
- Research data
- Competition datasets
- Transactional data



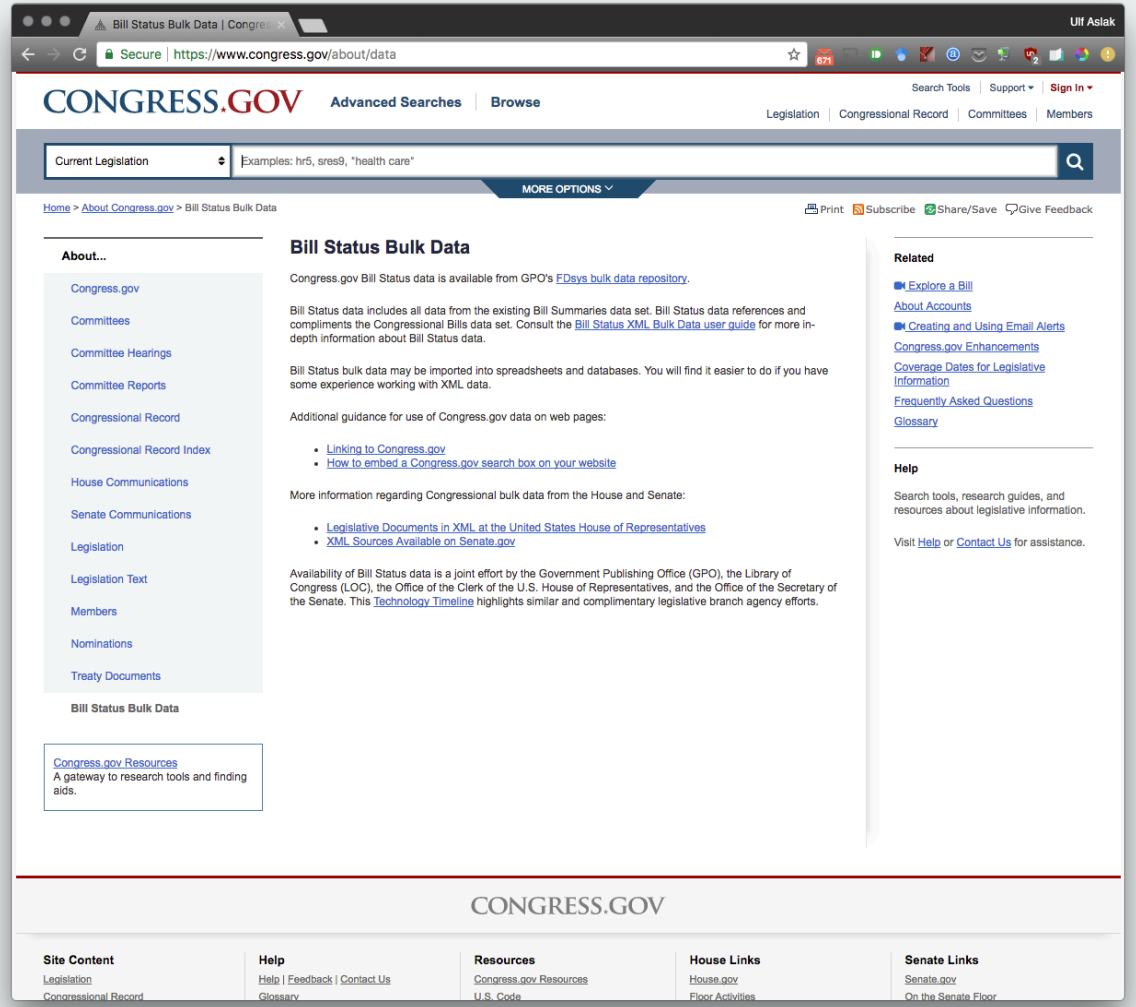
Open data

- Geographical data
- City data
- **Political data**
- Research data
- Competition datasets
- Transactional data



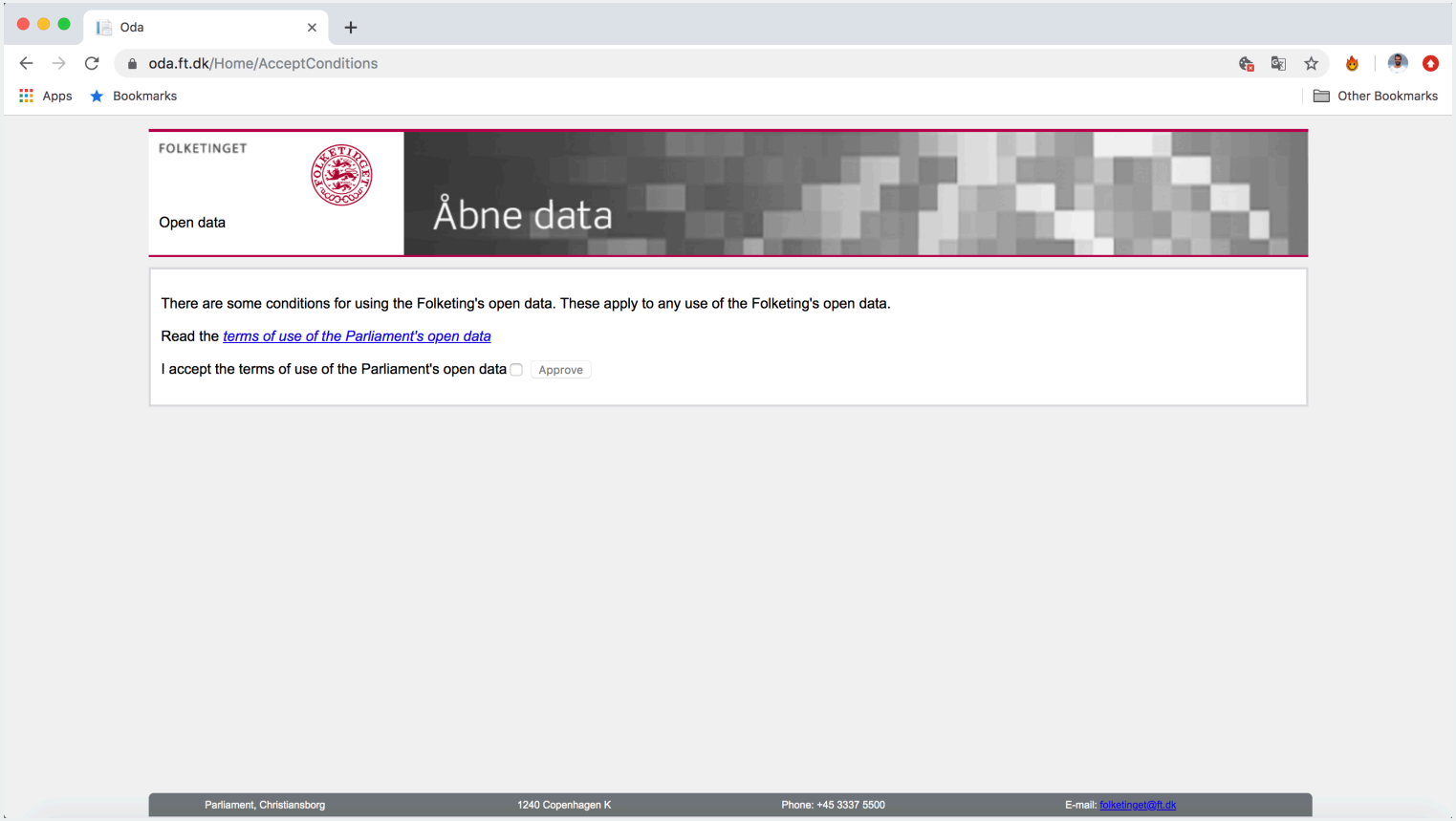
Open data

- Geographical data
- City data
- **Political data**
- Research data
- Competition datasets
- Transactional data



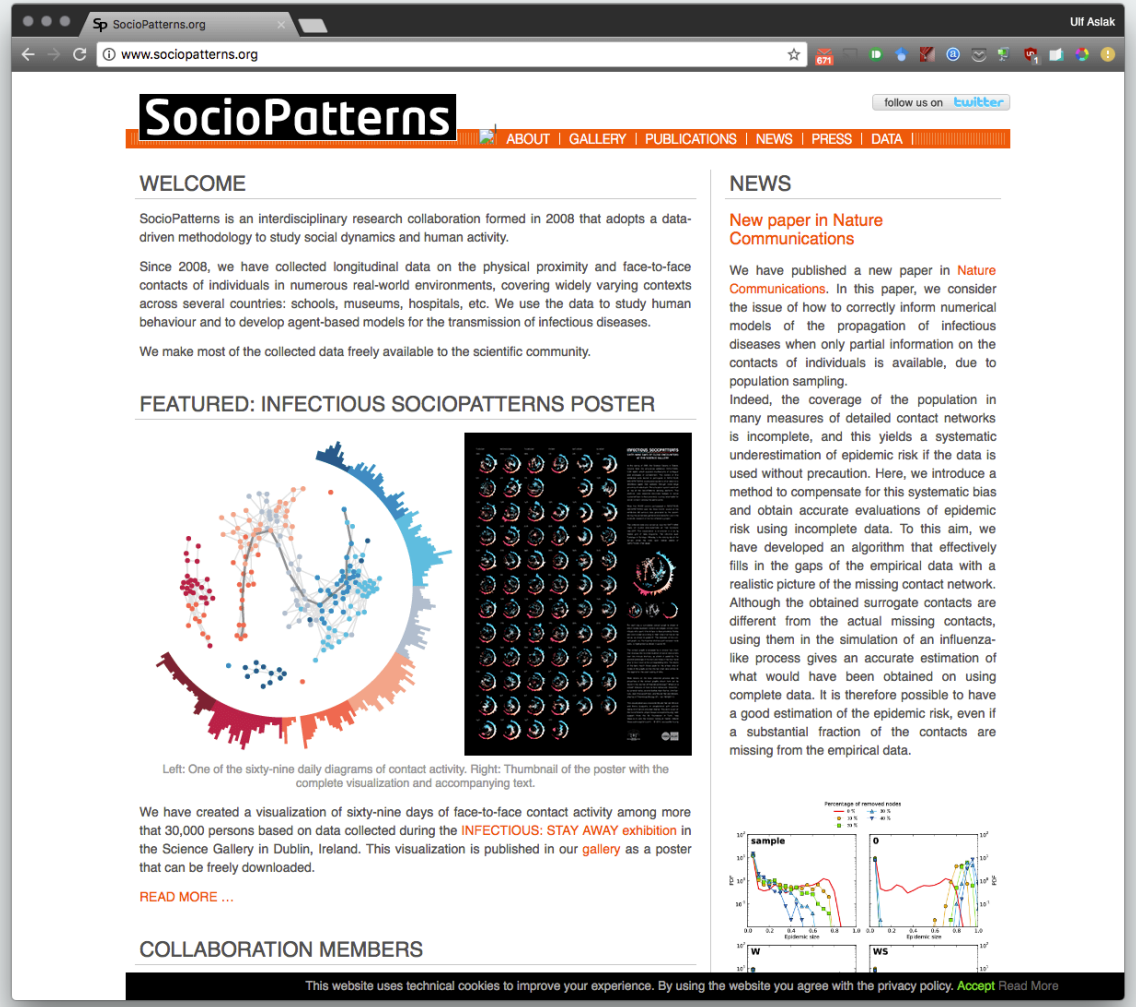
Open data

- Geographical data
- City data
- **Political data**
- Research data
- Competition datasets
- Transactional data



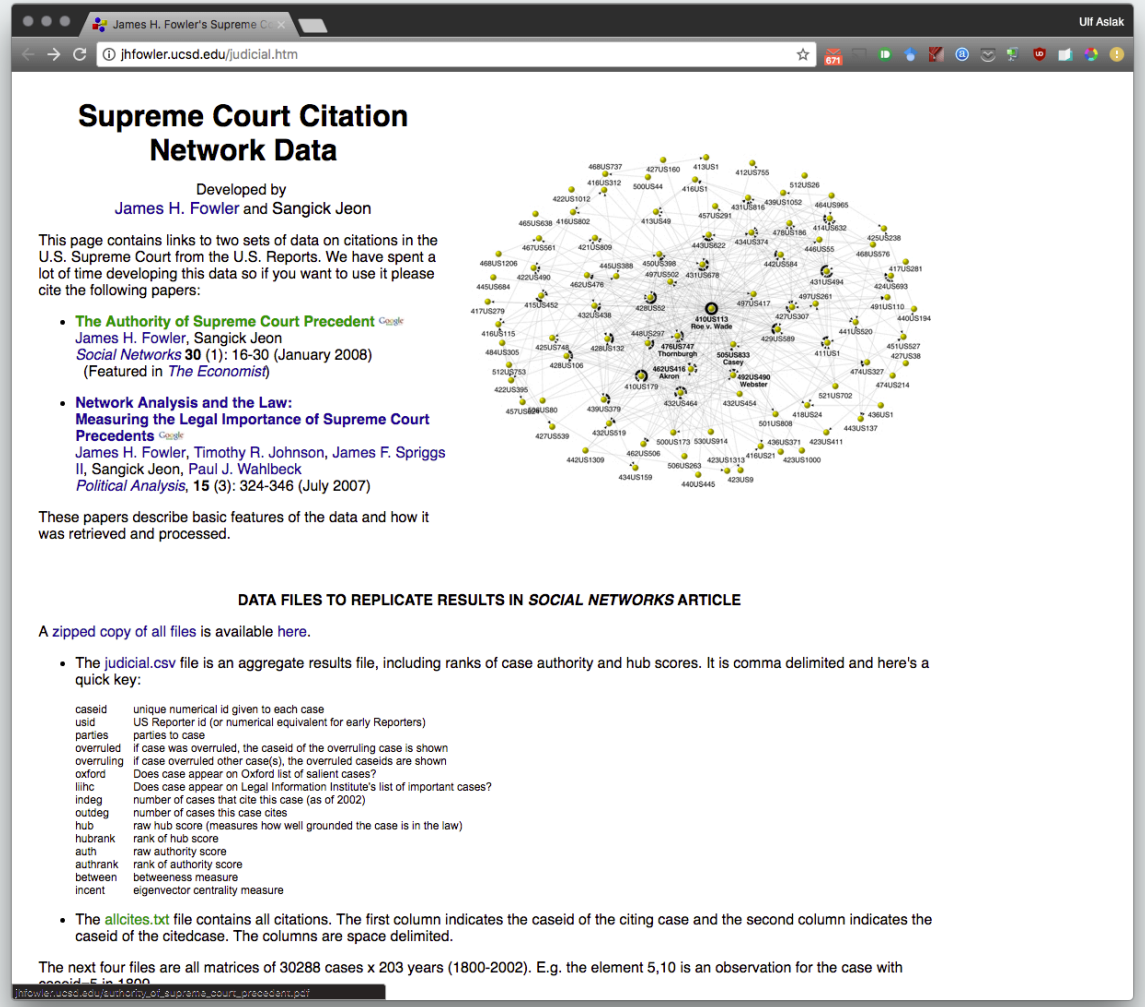
Open data

- Geographical data
- City data
- Political data
- **Research data**
- Competition datasets
- Transactional data



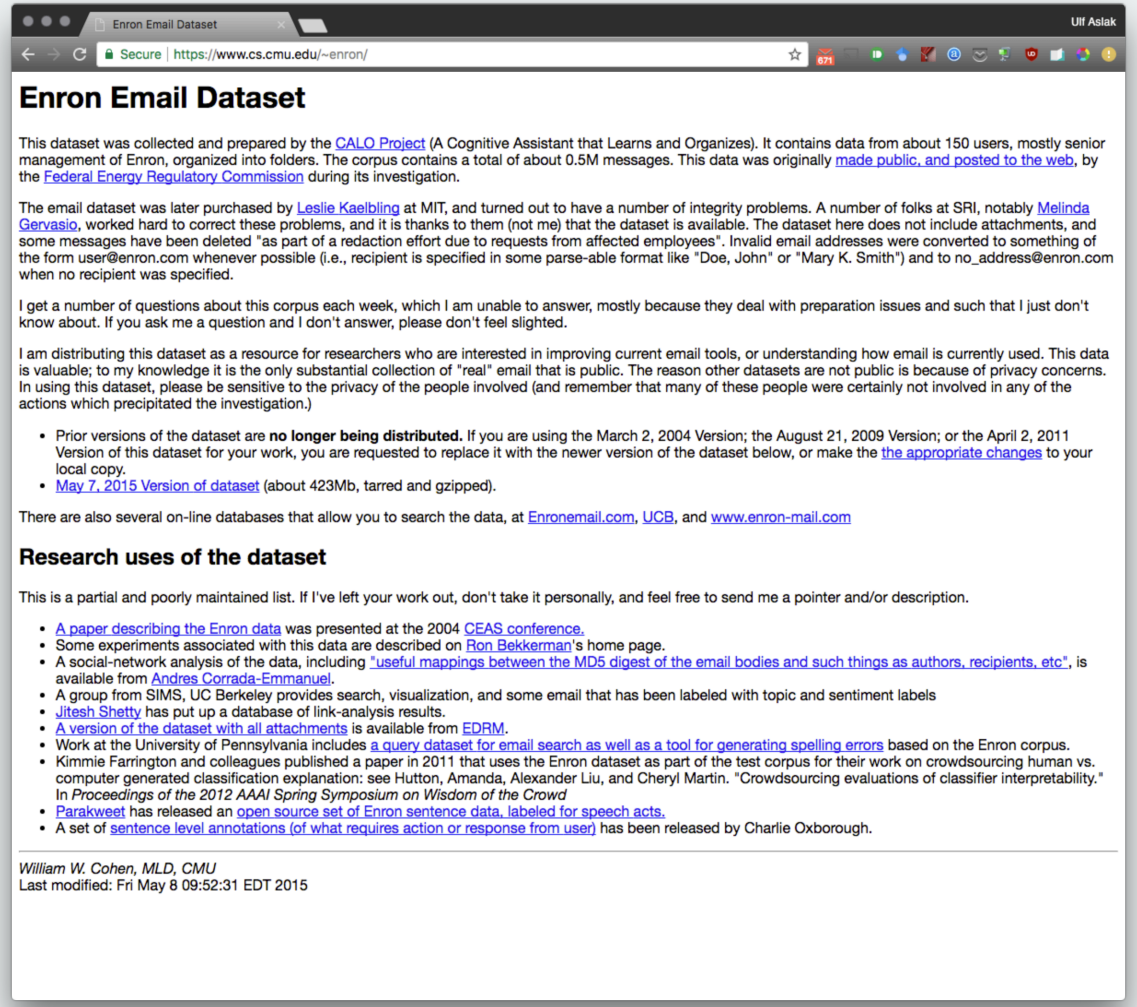
Open data

- Geographical data
- City data
- Political data
- **Research data**
- Competition datasets
- Transactional data



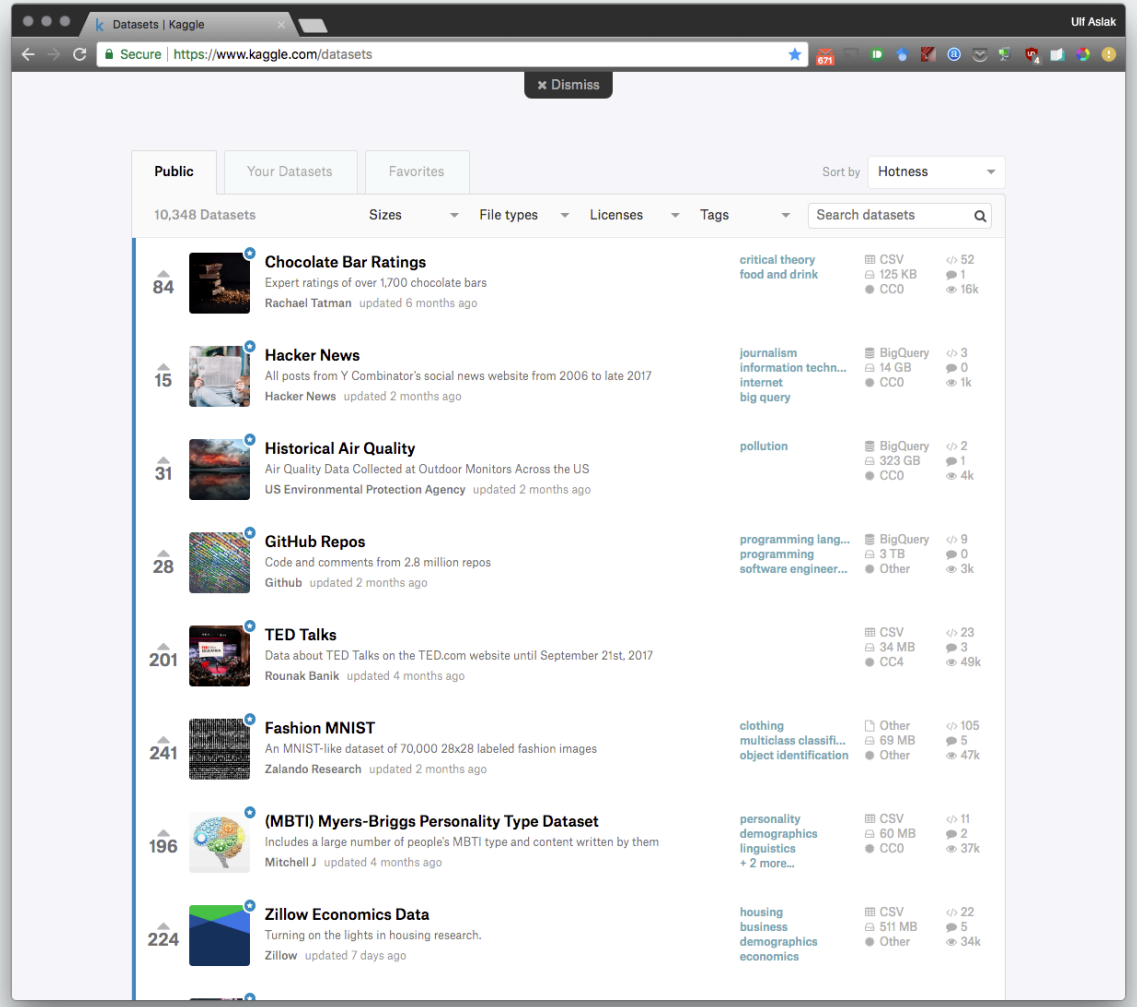
Open data

- Geographical data
- City data
- Political data
- **Research data**
- Competition datasets
- Transactional data



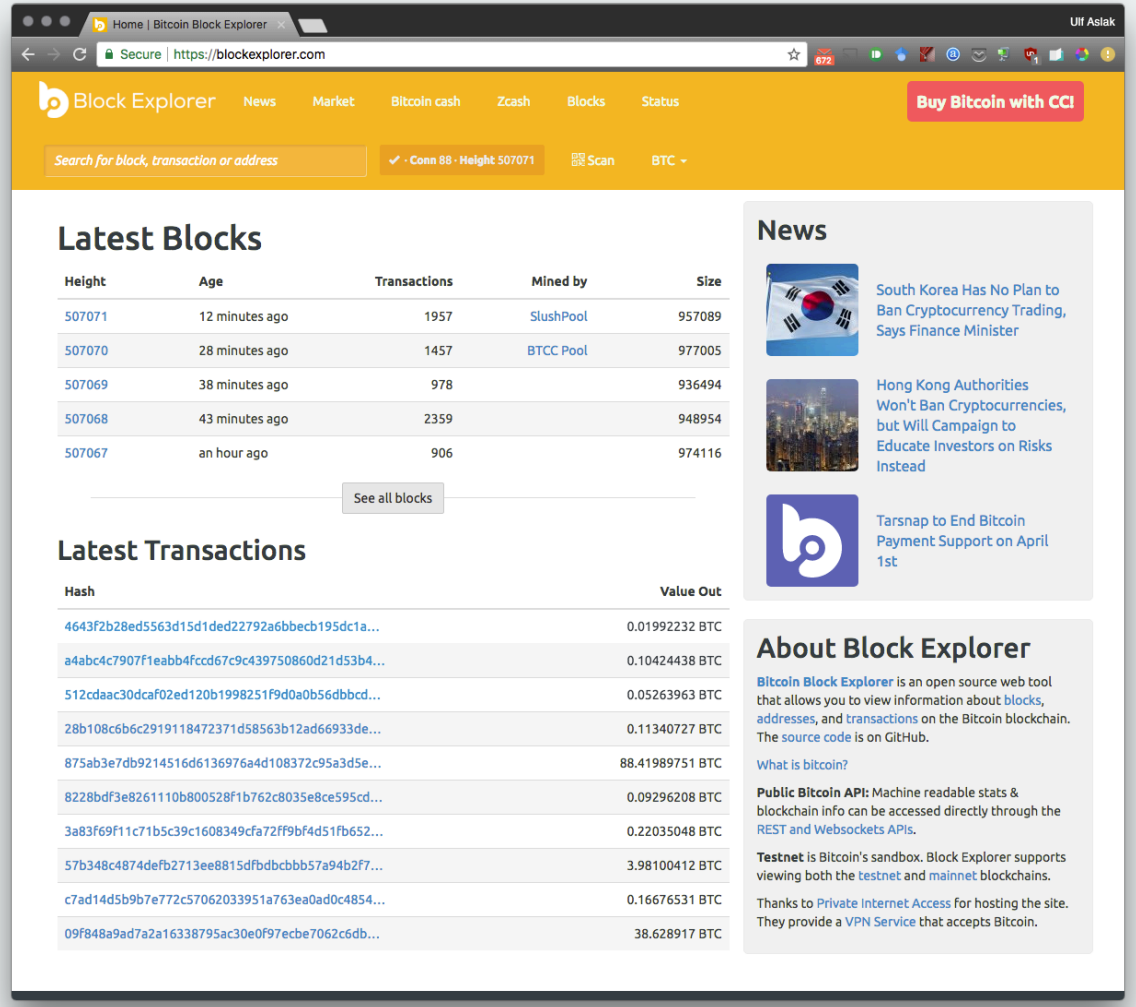
Open data

- Geographical data
- City data
- Political data
- Research data
- **Competition datasets**
- Transactional data



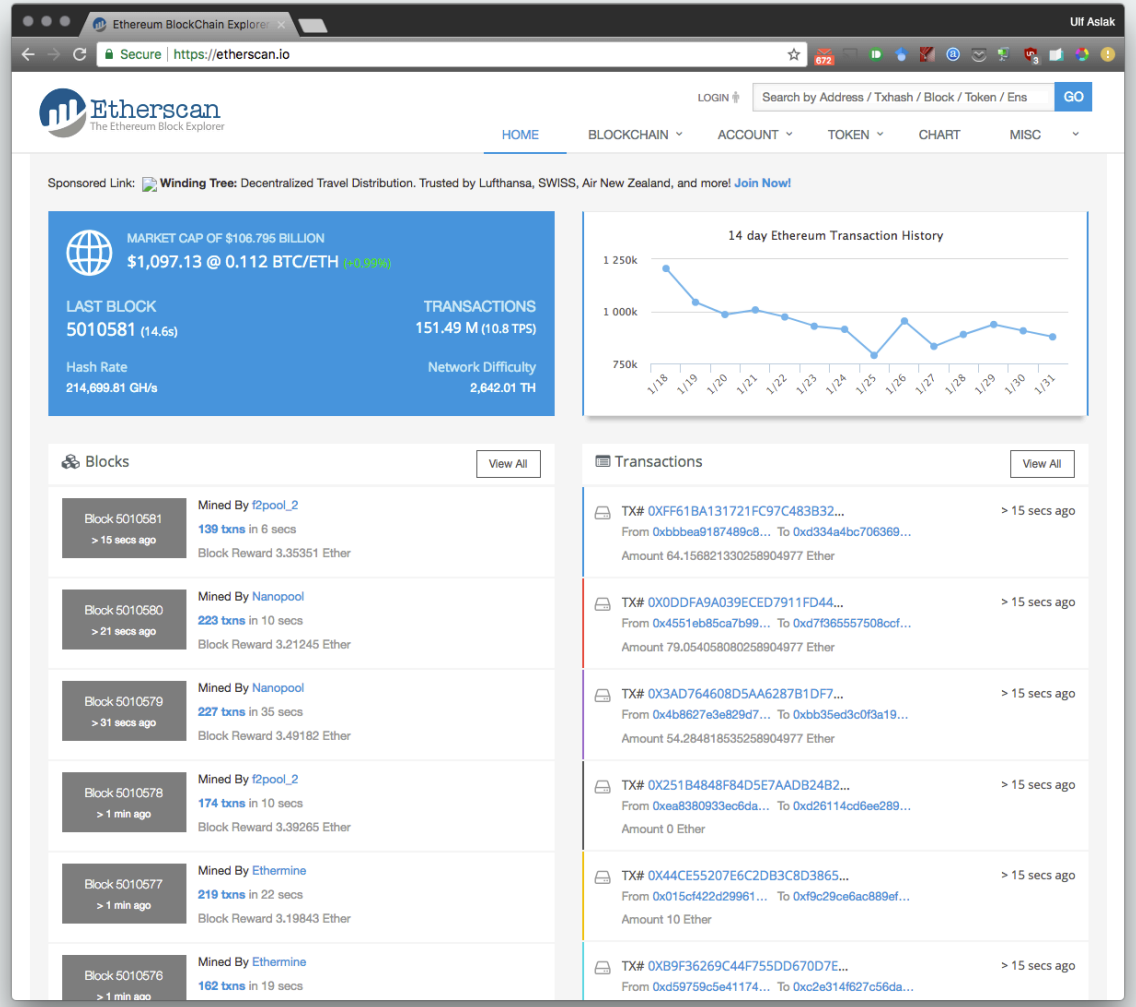
Open data

- Geographical data
- City data
- Political data
- Research data
- Competition datasets
- **Transactional data**



Open data

- Geographical data
- City data
- Political data
- Research data
- Competition datasets
- **Transactional data**

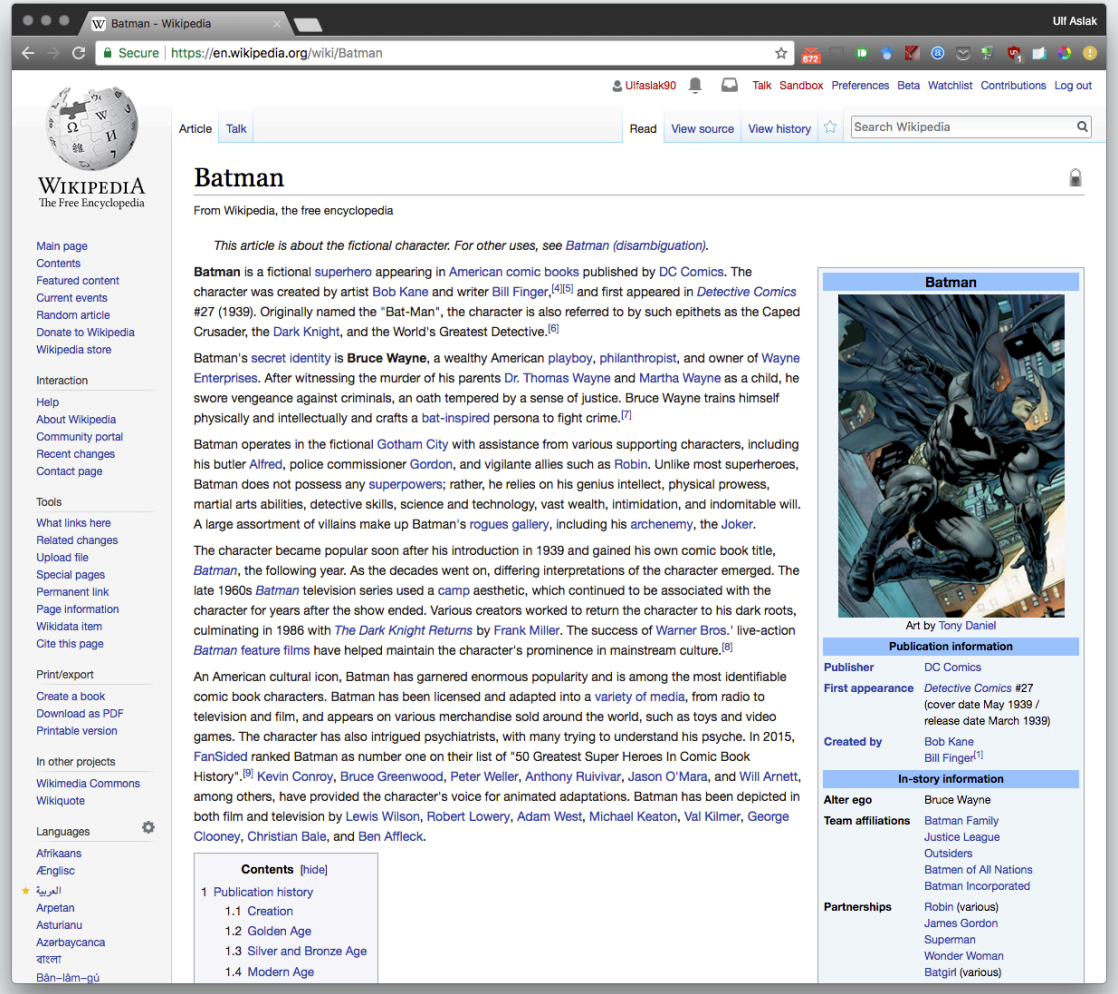


APIs

APIs

What is an API?

APIs

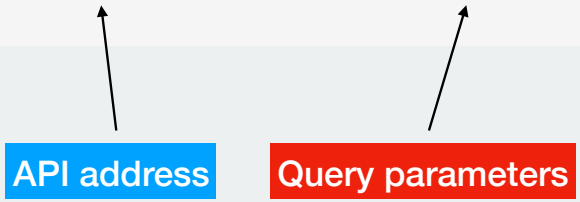

Batman - Wikipedia
Ulf Aslak
Secure | https://en.wikipedia.org/wiki/Batman
Ulfaslak90 Talk Sandbox Preferences Beta Watchlist Contributions Log out
Article Talk Read View source View history Search Wikipedia
WIKIPEDIA
The Free Encyclopedia
Batman
From Wikipedia, the free encyclopedia
This article is about the fictional character. For other uses, see Batman (disambiguation).
Batman is a fictional superhero appearing in American comic books published by DC Comics. The character was created by artist Bob Kane and writer Bill Finger,[4][5] and first appeared in Detective Comics #27 (1939). Originally named the "Bat-Man", the character is also referred to by such epithets as the Caped Crusader, the Dark Knight, and the World's Greatest Detective.[6]
Batman's secret identity is Bruce Wayne, a wealthy American playboy, philanthropist, and owner of Wayne Enterprises. After witnessing the murder of his parents Dr. Thomas Wayne and Martha Wayne as a child, he swore vengeance against criminals, an oath tempered by a sense of justice. Bruce Wayne trains himself physically and intellectually and crafts a bat-inspired persona to fight crime.[7]
Batman operates in the fictional Gotham City with assistance from various supporting characters, including his butler Alfred, police commissioner Gordon, and vigilante allies such as Robin. Unlike most superheroes, Batman does not possess any superpowers; rather, he relies on his genius intellect, physical prowess, martial arts abilities, detective skills, science and technology, vast wealth, intimidation, and indomitable will. A large assortment of villains make up Batman's rogues gallery, including his archenemy, the Joker.
The character became popular soon after his introduction in 1939 and gained his own comic book title, Batman, the following year. As the decades went on, differing interpretations of the character emerged. The late 1960s Batman television series used a camp aesthetic, which continued to be associated with the character for years after the show ended. Various creators worked to return the character to his dark roots, culminating in 1986 with The Dark Knight Returns by Frank Miller. The success of Warner Bros.' live-action Batman feature films have helped maintain the character's prominence in mainstream culture.[8]
An American cultural icon, Batman has garnered enormous popularity and is among the most identifiable comic book characters. Batman has been licensed and adapted into a variety of media, from radio to television and film, and appears on various merchandise sold around the world, such as toys and video games. The character has also intrigued psychiatrists, with many trying to understand his psyche. In 2015, FanSided ranked Batman as number one on their list of "50 Greatest Super Heroes In Comic Book History".[9] Kevin Conroy, Bruce Greenwood, Peter Weller, Anthony Ruivivar, Jason O'Mara, and Will Arnett, among others, have provided the character's voice for animated adaptations. Batman has been depicted in both film and television by Lewis Wilson, Robert Lowery, Adam West, Michael Keaton, Val Kilmer, George Clooney, Christian Bale, and Ben Affleck.
Contents [hide]
1 Publication history
1.1 Creation
1.2 Golden Age
1.3 Silver and Bronze Age
1.4 Modern Age
Batman
Art by Tony Daniel
Publication information
Publisher DC Comics
First appearance Detective Comics #27 (cover date May 1939 / release date March 1939)
Created by Bob Kane Bill Finger[1]
In-story information
Alter ego Bruce Wayne
Team affiliations Batman Family Justice League Outsiders Batmen of All Nations Batman Incorporated
Partnerships Robin (various) James Gordon Superman Wonder Woman Batgirl (various)

APIs

```
1 import requests as rq
2
3 query = "https://en.wikipedia.org/w/api.php?format=json&action=query&titles=Batman&prop=revisions&rvprop=content"
4 response = rq.get(query)
5 result = response.json()
```

APIs

```
1 import requests as rq
2
3 query = "https://en.wikipedia.org/w/api.php?format=json&action=query&titles=Batman&prop=revisions&rvprop=content"
4 response = rq.get(query)
5 result = response.json()
```



APIs

```
1 import requests as rq
2
3 query = "https://en.wikipedia.org/w/api.php?format=json&action=query&titles=Batman&prop=revisions&rvprop=content"
4 response = rq.get(query)
5 result = response.json()
```

APIs

```
1 import requests as rq
2
3 query = "https://en.wikipedia.org/w/api.php?format=json&action=query&titles=Batman&prop=revisions&rvprop=content"
4 response = rq.get(query)
5 result = response.json()
```

APIs

```
1 import requests as rq
2
3 query = "https://en.wikipedia.org/w/api.php?format=json&action=query&titles=Batman&prop=revisions&rvprop=content"
4 response = rq.get(query)
5 result = response.json()
```

APIs

```python
import requests as rq

query = "https://en.wikipedia.org/w/api.php?format=json&action=query&titles=Batman&prop=revisions&rvprop=content"
response = rq.get(query)
result = response.json()
```

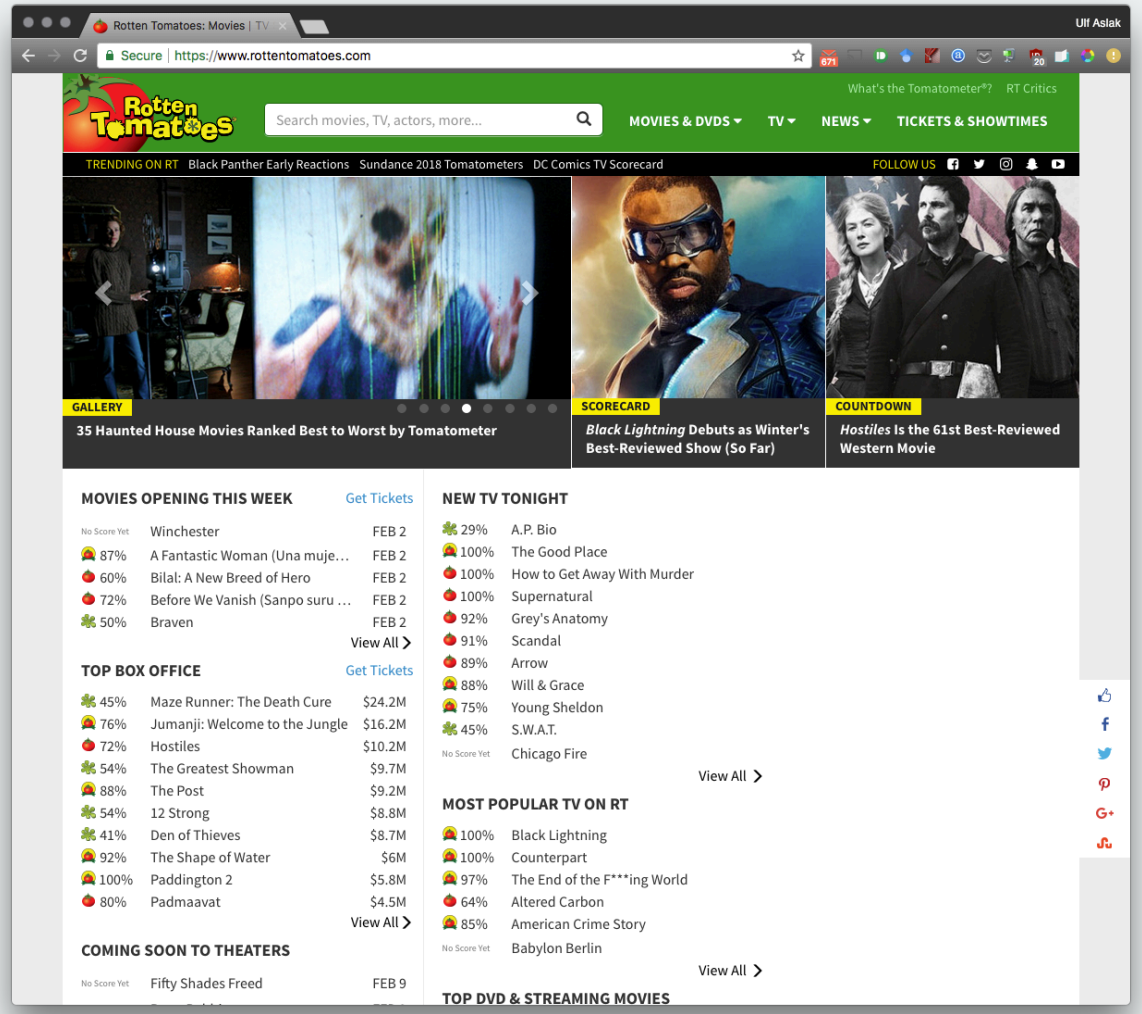
```python
import json

print json.dumps(result, indent=4)
```

Last executed 2018-02-01 10:28:27 in 10ms

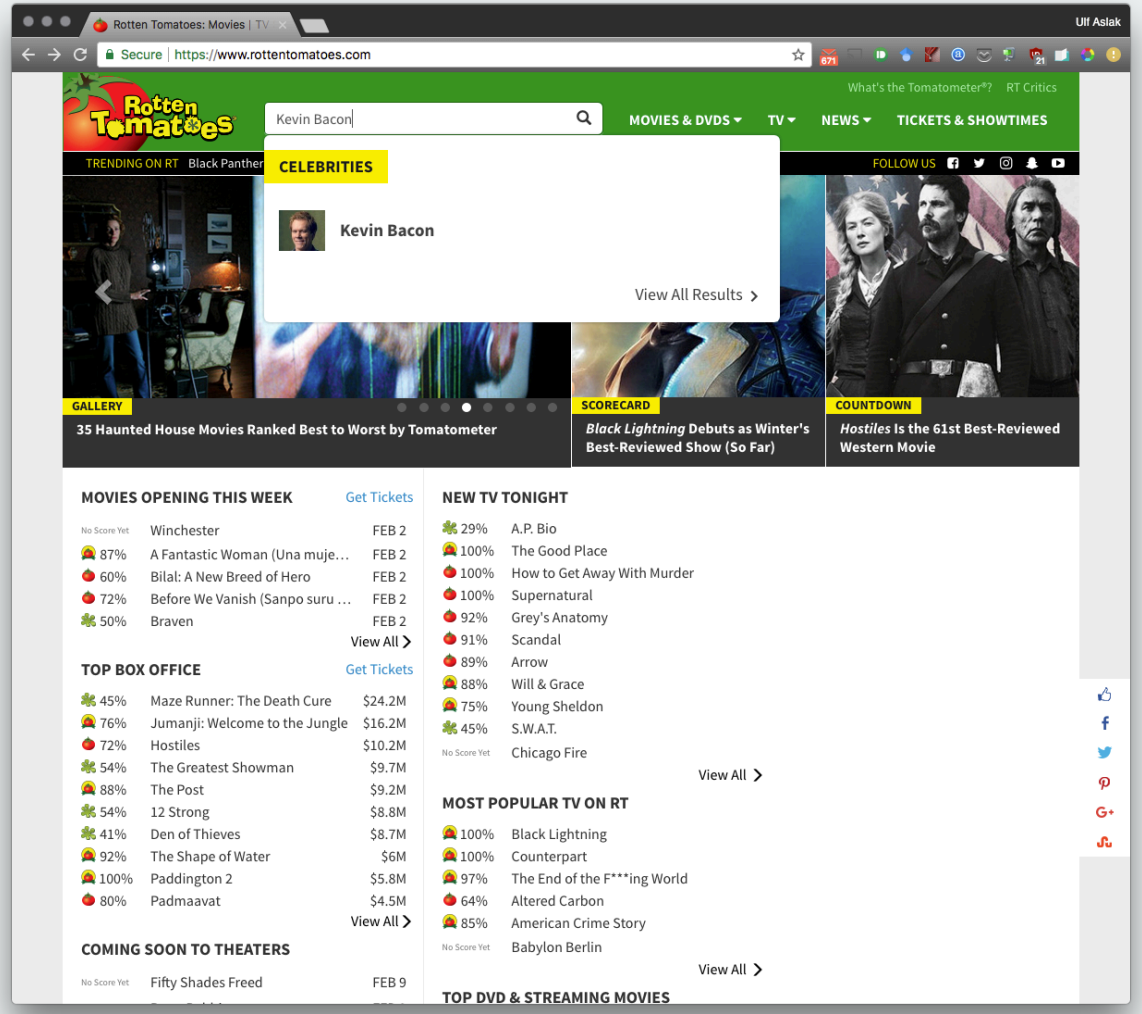

```
{
    "batchcomplete": "",
    "query": {
        "pages": {
            "4335": {
                "ns": 0,
                "pageid": 4335,
                "revisions": [
                    {
                        "*": "{{About|the fictional character}}\n{{pp-semi-indef}}\n{{pp-move-indef}}\n{{Use mdy da
tes|date=October 2015}}\n{{Infobox comics character\n|image          = Batman Detective Comics Vol 2 1.png<!--Do NO
T change this image without consensus from the Talk Page-->\n|imagesize      =\n|converted      = y\n|caption
  = Art by [[Tony Daniel]]\n|alt            = Batman descends upon Gotham City\n|publisher      = [[DC Comics]]\n|d
ebut          = ''[[Detective Comics]]'' #27<br />(cover date May 1939 /<br>release date March 1939)<!-- \"Debut\"
indicates the first appearance of a character, not a change to the character's backstory. -->\n|creators       =
{{plainlist|\n*[[Bob Kane]]\n*[[Bill Finger]]<ref>[[Ron Goulart|Goulart, Ron]], ''Comic Book Encyclopedia'' ([[Harp
erCollins|Harper Entertainment]], New York, 2004) {{ISBN|978-0-06-053816-3}}</ref>\n}}\n|alter_ego      = Bruce Way
ne<!-- Do not enter a middle name. He has been depicted with too many different middle names to enter a specific on
e here. Also, there is no past or current, dead, or alive in fiction from a real world perspective; the infobox sho
uld cover the Batman known to the public consciousness and not a current comic book storyline. -->\n|alliances
= {{plainlist|\n*[[List of Batman supporting characters#Batman Family|Batman Family]]\n*[[Justice League]]\n*[[Outs
```

(web) Scraping

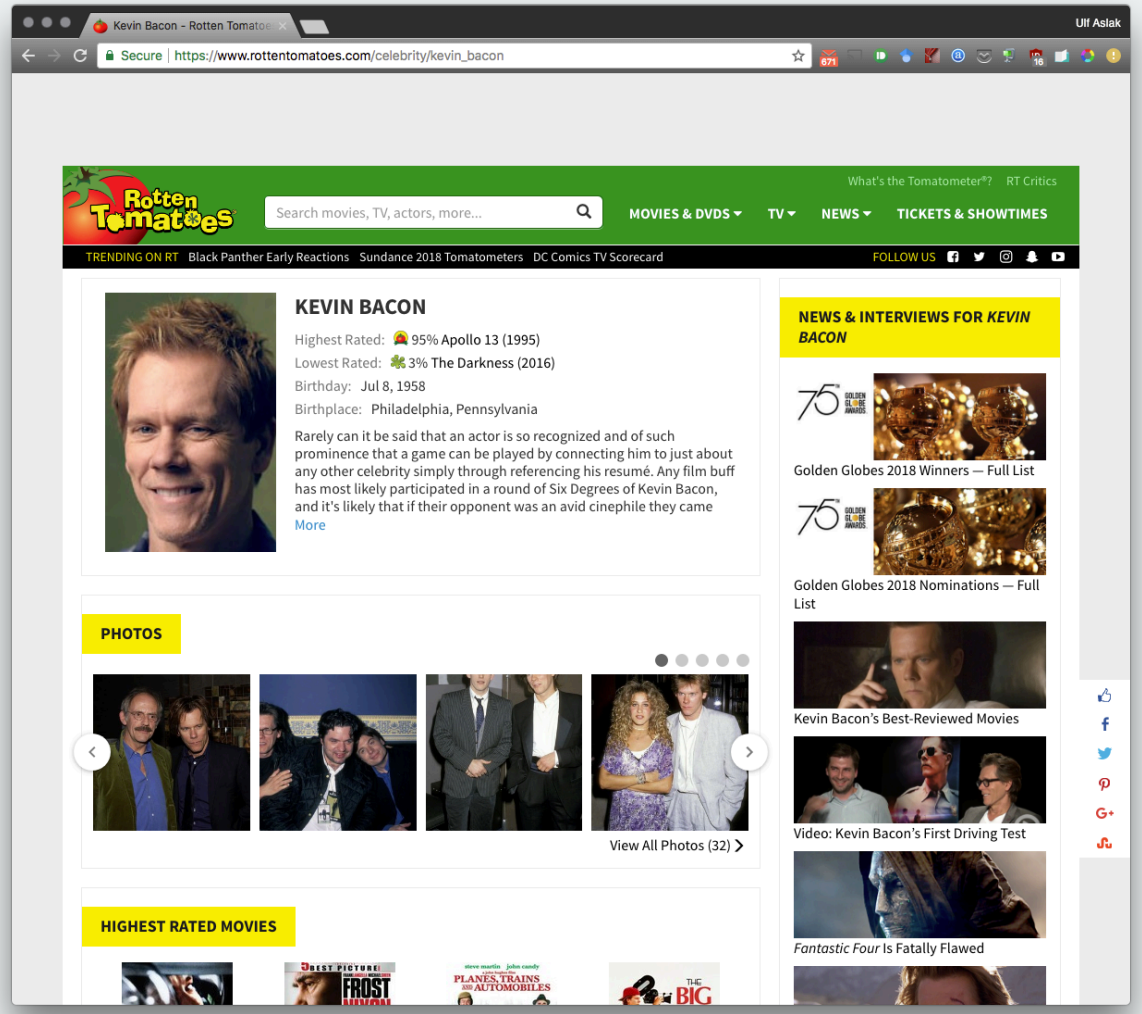
Scraping


Rotten Tomatoes: Movies | TV
Ulf Aslak
Secure | https://www.rottentomatoes.com
What's the Tomatometer®? RT Critics
Search movies, TV, actors, more...
MOVIES & DVDS
TV
NEWS
TICKETS & SHOWTIMES
TRENDING ON RT Black Panther Early Reactions Sundance 2018 Tomatometers DC Comics TV Scorecard
FOLLOW US
GALLERY
35 Haunted House Movies Ranked Best to Worst by Tomatometer
SCORECARD
Black Lightning Debuts as Winter's Best-Reviewed Show (So Far)
COUNTDOWN
Hostiles Is the 61st Best-Reviewed Western Movie
MOVIES OPENING THIS WEEK
Get Tickets
No Score Yet Winchester FEB 2
87% A Fantastic Woman (Una muje… FEB 2
60% Bilal: A New Breed of Hero FEB 2
72% Before We Vanish (Sanpo suru … FEB 2
50% Braven FEB 2
View All
TOP BOX OFFICE
Get Tickets
45% Maze Runner: The Death Cure $24.2M
76% Jumanji: Welcome to the Jungle $16.2M
72% Hostiles $10.2M
54% The Greatest Showman $9.7M
88% The Post $9.2M
54% 12 Strong $8.8M
41% Den of Thieves $8.7M
92% The Shape of Water $6M
100% Paddington 2 $5.8M
80% Padmaavat $4.5M
View All
COMING SOON TO THEATERS
No Score Yet Fifty Shades Freed FEB 9
NEW TV TONIGHT
29% A.P. Bio
100% The Good Place
100% How to Get Away With Murder
100% Supernatural
92% Grey's Anatomy
91% Scandal
89% Arrow
88% Will & Grace
75% Young Sheldon
45% S.W.A.T.
No Score Yet Chicago Fire
View All
MOST POPULAR TV ON RT
100% Black Lightning
100% Counterpart
97% The End of the F***ing World
64% Altered Carbon
85% American Crime Story
No Score Yet Babylon Berlin
View All
TOP DVD & STREAMING MOVIES

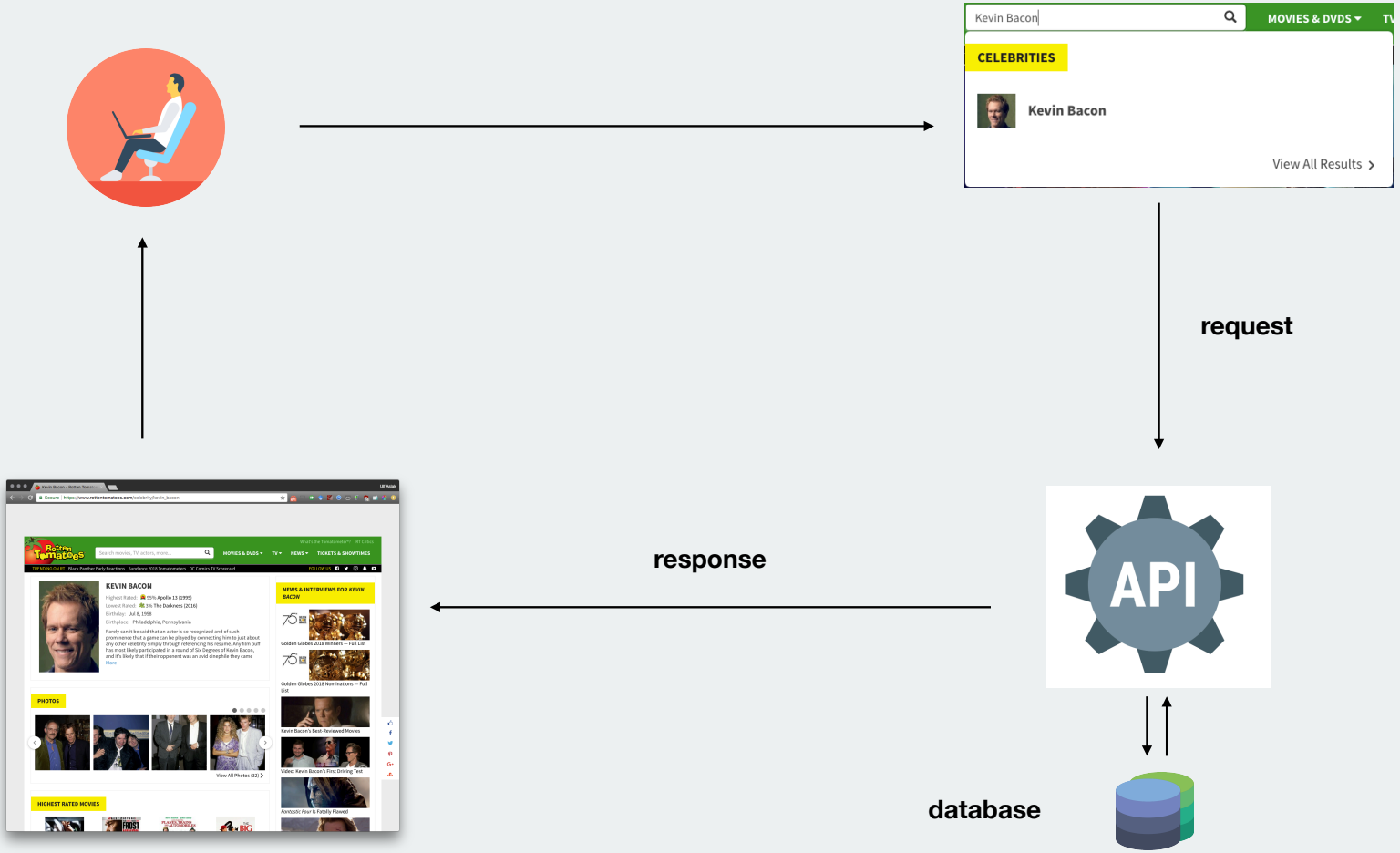
Scraping


Rotten Tomatoes: Movies | TV
Ulf Aslak
Secure | https://www.rottentomatoes.com
What's the Tomatometer®? RT Critics
Kevin Bacon
MOVIES & DVDS
TV
NEWS
TICKETS & SHOWTIMES
TRENDING ON RT Black Panther
FOLLOW US
CELEBRITIES
Kevin Bacon
View All Results
GALLERY
35 Haunted House Movies Ranked Best to Worst by Tomatometer
SCORECARD
Black Lightning Debuts as Winter's Best-Reviewed Show (So Far)
COUNTDOWN
Hostiles Is the 61st Best-Reviewed Western Movie
MOVIES OPENING THIS WEEK
Get Tickets
No Score Yet Winchester FEB 2
87% A Fantastic Woman (Una muje… FEB 2
60% Bilal: A New Breed of Hero FEB 2
72% Before We Vanish (Sanpo suru … FEB 2
50% Braven FEB 2
View All
TOP BOX OFFICE
Get Tickets
45% Maze Runner: The Death Cure $24.2M
76% Jumanji: Welcome to the Jungle $16.2M
72% Hostiles $10.2M
54% The Greatest Showman $9.7M
88% The Post $9.2M
54% 12 Strong $8.8M
41% Den of Thieves $8.7M
92% The Shape of Water $6M
100% Paddington 2 $5.8M
80% Padmaavat $4.5M
View All
COMING SOON TO THEATERS
No Score Yet Fifty Shades Freed FEB 9
NEW TV TONIGHT
29% A.P. Bio
100% The Good Place
100% How to Get Away With Murder
100% Supernatural
92% Grey's Anatomy
91% Scandal
89% Arrow
88% Will & Grace
75% Young Sheldon
45% S.W.A.T.
No Score Yet Chicago Fire
View All
MOST POPULAR TV ON RT
100% Black Lightning
100% Counterpart
97% The End of the F***ing World
64% Altered Carbon
85% American Crime Story
No Score Yet Babylon Berlin
View All
TOP DVD & STREAMING MOVIES

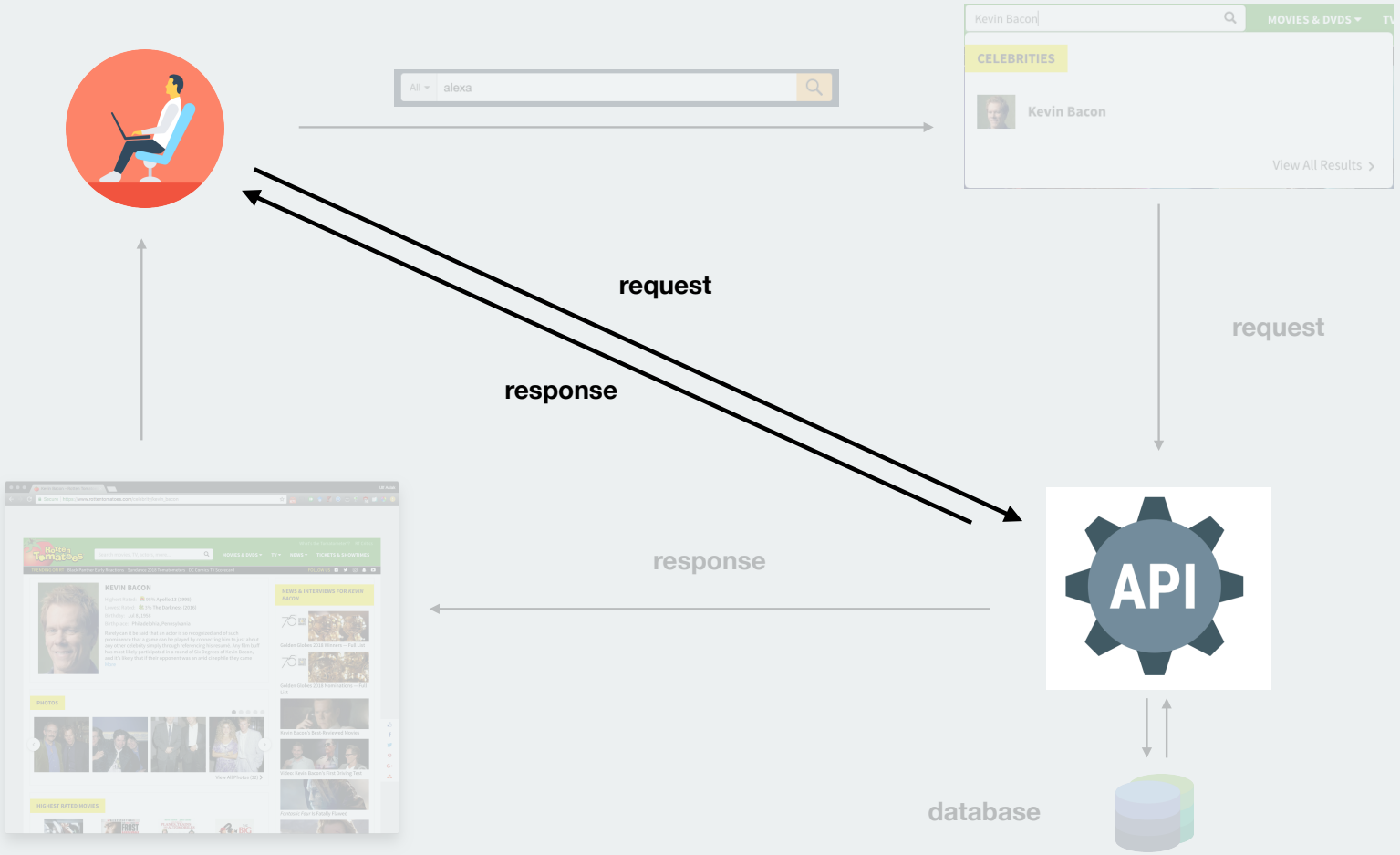
Scraping


Kevin Bacon - Rotten Tomatoes
Ulf Aslak
Secure | https://www.rottentomatoes.com/celebrity/kevin_bacon
Rotten Tomatoes
Search movies, TV, actors, more...
What's the Tomatometer®? RT Critics
MOVIES & DVDS
TV
NEWS
TICKETS & SHOWTIMES
TRENDING ON RT Black Panther Early Reactions Sundance 2018 Tomatometers DC Comics TV Scorecard
FOLLOW US
KEVIN BACON
Highest Rated: 95% Apollo 13 (1995)
Lowest Rated: 3% The Darkness (2016)
Birthday: Jul 8, 1958
Birthplace: Philadelphia, Pennsylvania
Rarely can it be said that an actor is so recognized and of such prominence that a game can be played by connecting him to just about any other celebrity simply through referencing his resumé. Any film buff has most likely participated in a round of Six Degrees of Kevin Bacon, and it's likely that if their opponent was an avid cinephile they came
More
PHOTOS
View All Photos (32)
HIGHEST RATED MOVIES
NEWS & INTERVIEWS FOR KEVIN BACON
Golden Globes 2018 Winners — Full List
Golden Globes 2018 Nominations — Full List
Kevin Bacon's Best-Reviewed Movies
Video: Kevin Bacon's First Driving Test
Fantastic Four Is Fatally Flawed

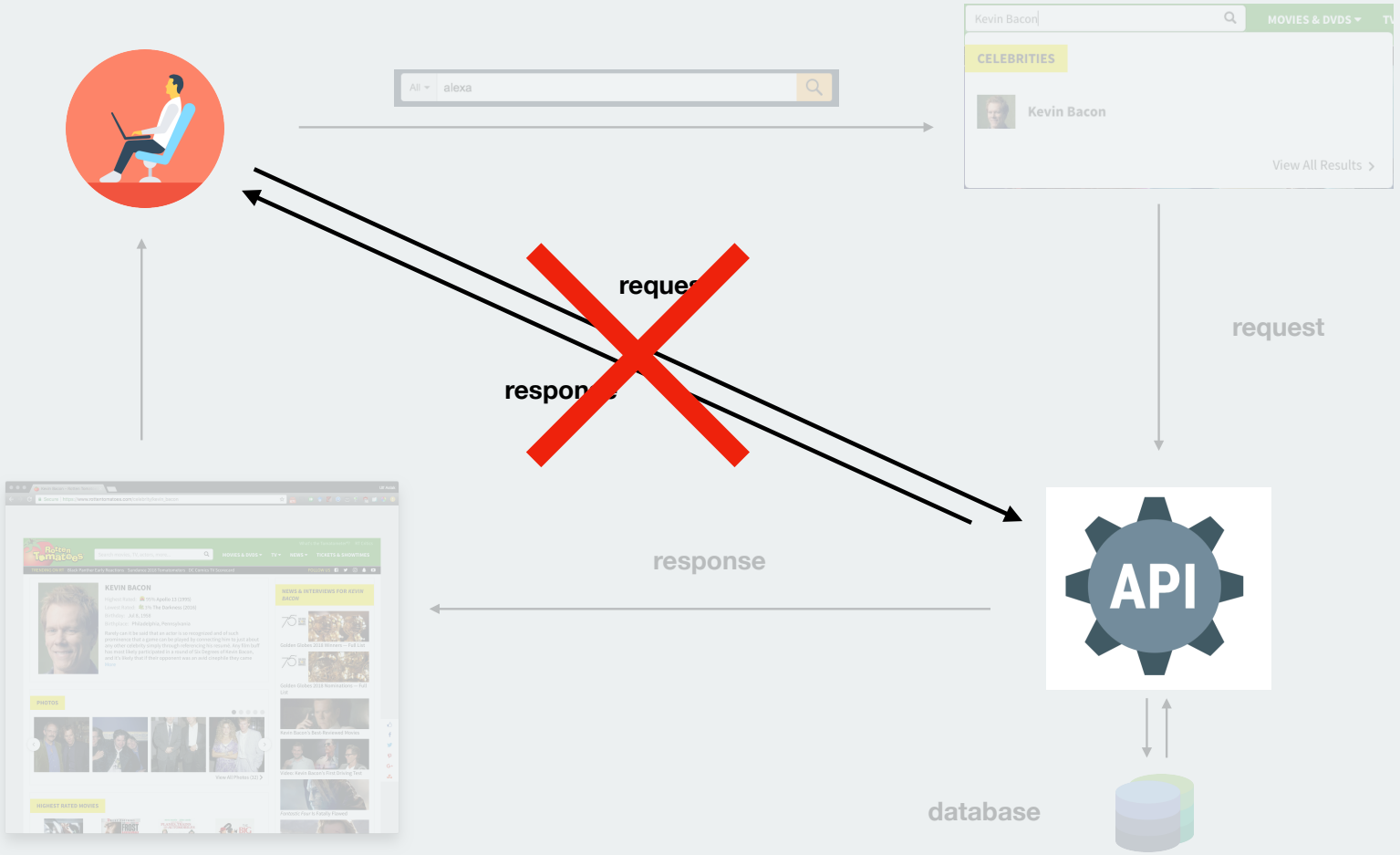
Scraping


Kevin Bacon
MOVIES & DVDS
CELEBRITIES
Kevin Bacon
View All Results
request
response
API
database

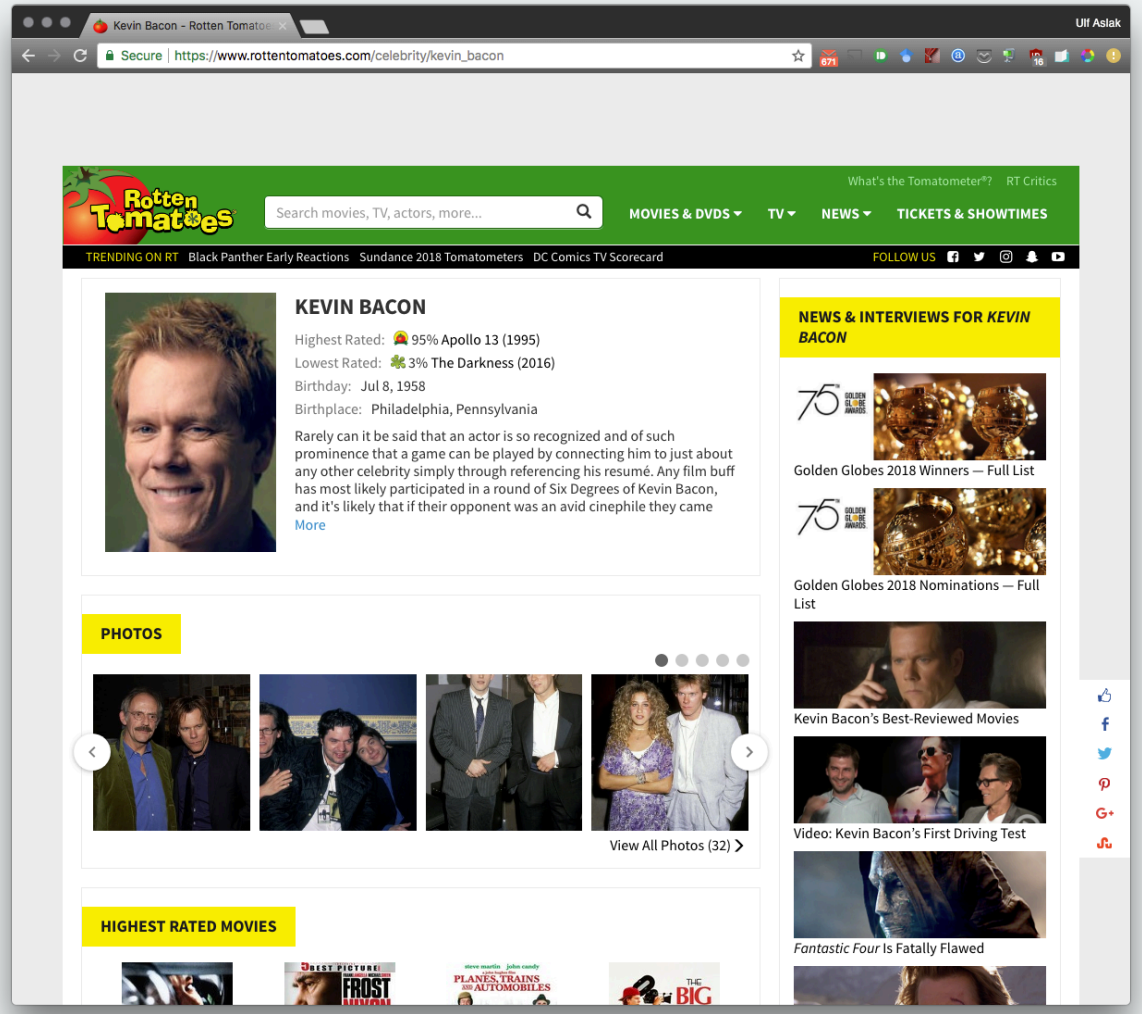
Scraping


alexa
Kevin Bacon
MOVIES & DVDS
CELEBRITIES
Kevin Bacon
View All Results
request
response
request
response
API
database

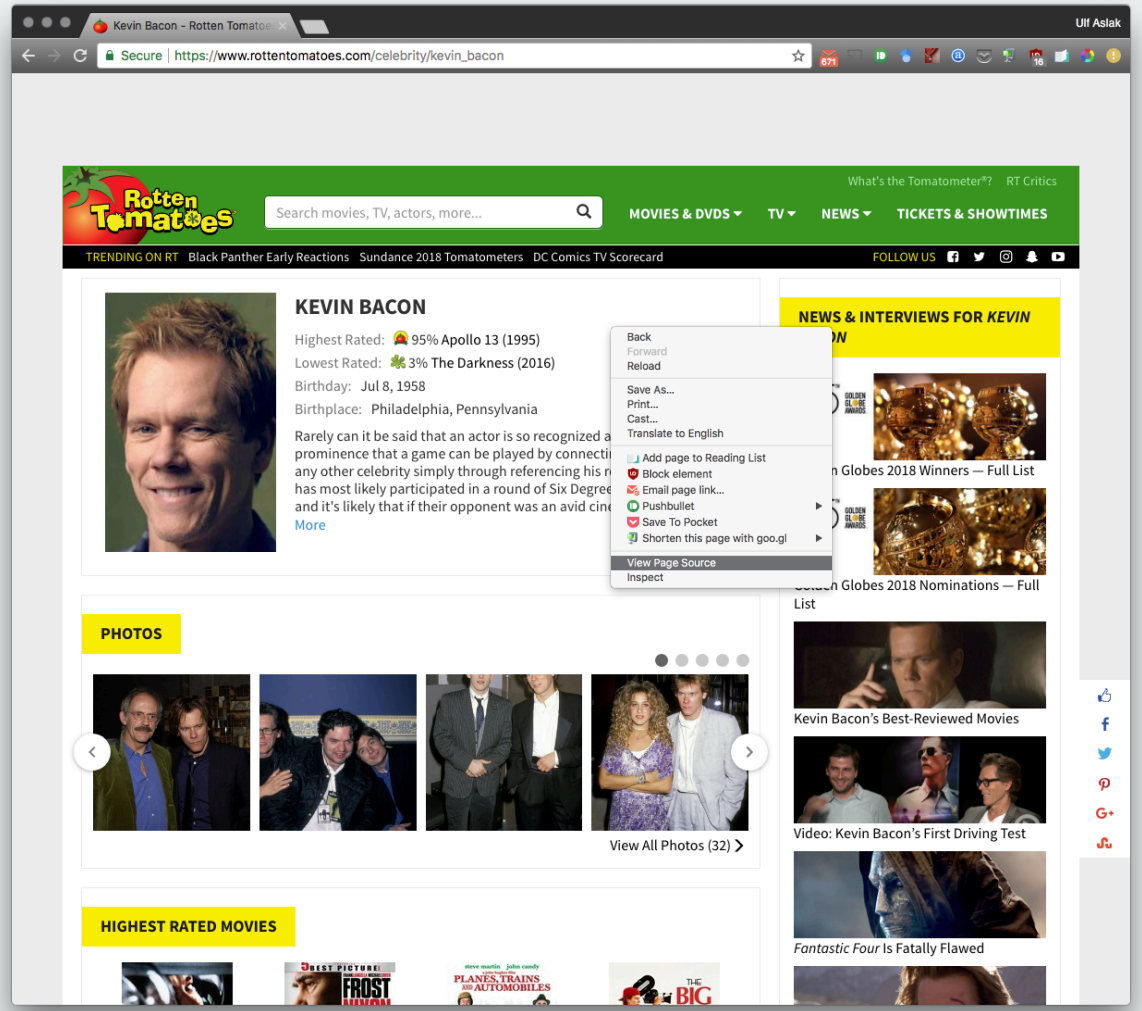
Scraping


alexa
Kevin Bacon
MOVIES & DVDS
CELEBRITIES
Kevin Bacon
View All Results
request
response
API
database

Scraping


Kevin Bacon - Rotten Tomatoes
Ulf Aslak
Secure | https://www.rottentomatoes.com/celebrity/kevin_bacon
Rotten Tomatoes
Search movies, TV, actors, more...
What's the Tomatometer®? RT Critics
MOVIES & DVDS
TV
NEWS
TICKETS & SHOWTIMES
TRENDING ON RT Black Panther Early Reactions Sundance 2018 Tomatometers DC Comics TV Scorecard
FOLLOW US
KEVIN BACON
Highest Rated: 95% Apollo 13 (1995)
Lowest Rated: 3% The Darkness (2016)
Birthday: Jul 8, 1958
Birthplace: Philadelphia, Pennsylvania
Rarely can it be said that an actor is so recognized and of such prominence that a game can be played by connecting him to just about any other celebrity simply through referencing his resumé. Any film buff has most likely participated in a round of Six Degrees of Kevin Bacon, and it's likely that if their opponent was an avid cinephile they came
More
PHOTOS
View All Photos (32)
HIGHEST RATED MOVIES
NEWS & INTERVIEWS FOR KEVIN BACON
Golden Globes 2018 Winners — Full List
Golden Globes 2018 Nominations — Full List
Kevin Bacon's Best-Reviewed Movies
Video: Kevin Bacon's First Driving Test
Fantastic Four Is Fatally Flawed

Scraping


Kevin Bacon - Rotten Tomatoes
Ulf Aslak
Secure | https://www.rottentomatoes.com/celebrity/kevin_bacon
Rotten Tomatoes
Search movies, TV, actors, more...
What's the Tomatometer®? RT Critics
MOVIES & DVDS
TV
NEWS
TICKETS & SHOWTIMES
TRENDING ON RT Black Panther Early Reactions Sundance 2018 Tomatometers DC Comics TV Scorecard
FOLLOW US
KEVIN BACON
Highest Rated: 95% Apollo 13 (1995)
Lowest Rated: 3% The Darkness (2016)
Birthday: Jul 8, 1958
Birthplace: Philadelphia, Pennsylvania
Rarely can it be said that an actor is so recognized a
prominence that a game can be played by connecti
any other celebrity simply through referencing his r
has most likely participated in a round of Six Degree
and it's likely that if their opponent was an avid cine
More
Back
Forward
Reload
Save As...
Print...
Cast...
Translate to English
Add page to Reading List
Block element
Email page link...
Pushbullet
Save To Pocket
Shorten this page with goo.gl
View Page Source
Inspect
NEWS & INTERVIEWS FOR KEVIN
Globes 2018 Winners — Full List
Globes 2018 Nominations — Full List
Kevin Bacon's Best-Reviewed Movies
Video: Kevin Bacon's First Driving Test
Fantastic Four Is Fatally Flawed
PHOTOS
View All Photos (32)
HIGHEST RATED MOVIES
FROST
PLANES, TRAINS AND AUTOMOBILES
THE BIG

Scraping

Always check GitHub:
probably someone has already written the scrapper you need

Recap

- Get **open data** from public institutions, researchers or data sharing sites.
- Request it from someone's **API**. Is very easy, but usually has limits.
- "Forcefully" take it by **scraping** it from a website

Get today's exercise from the GitHub repo!

Assignment 1 will come out at the end of this class
I will send an announcement about it :)