

# Utilizing APIs for Humanities Research

**Examples from The New York Times and The Guardian**

**Vincent J. Schober**

**13.12.2023**

# Overview

- API in General
- API Call/Response
- Data Structure and Conversion
- APIs in academia – why?
- New York Times API
- Guardian API

# Prerequisites to Working with APIs

- Some understanding of data structures necessary
- Some programming interest/experience is helpful

# What is an API?

- Application Programming Interface
- => basically a communication interface
  - With a specific set of rules



# The „Call“

- Communicating with an API is done through an API call or request
- The basic building blocs:
  - URL
  - Authorization
  - Paramterers
- Example:

<https://api.nytimes.com/svc/search/v2/articlesearch.json> ?q=election & api-key=yourkey

The diagram shows the URL `https://api.nytimes.com/svc/search/v2/articlesearch.json ?q=election & api-key=yourkey` with three blue brackets underneath. The first bracket spans the entire URL and is labeled "url". The second bracket spans the query string `?q=election &` and is labeled "params". The third bracket spans `api-key=yourkey` and is labeled "authorization".

- Tools: postman

# The „Response“

```
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28  
"status": "OK",  
"copyright": "Copyright (c) 2023 The New York Times Company. All Rights Reserved.",  
"response": {  
  "docs": [  
    {  
      "abstract": "The little-known former software executive had hoped his business acumen and relentless focus on  
        didn't.",  
      "web_url": "https://www.nytimes.com/2023/12/04/us/politics/doug-burgum-2024-campaign.html",  
      "snippet": "The little-known former software executive had hoped his business acumen and relentless focus on t  
        didn't.",  
      "lead_paragraph": "Gov. Doug Burgum of North Dakota, the wealthy former software executive who entered the pre  
        economy would propel him forward, dropped out of the race for the Republican nomination on Monday.",  
      "source": "The New York Times",  
      "multimedia": [  
        {  
          "rank": 0,  
          "subtype": "xlarge",  
          "caption": null,  
          "credit": null,  
          "type": "image",  
          "url": "images/2023/10/11/us/politics/00pol-BurgumHF0/00pol-BurgumHF0-articleLarge.jpg",  
          "height": 400,  
          "width": 600,  
          "legacy": {  
            "xlarge": "images/2023/10/11/us/politics/00pol-BurgumHF0/00pol-BurgumHF0-articleLarge.jpg",  
            "xlargewidth": 600,  
            "xlargeheight": 400  
          },  
          "subType": "xlarge",  
          "crop_name": "articleLarge"  
        }  
      ]  
    }  
  ]  
}
```



# The „Response“

- JSON – Java Script Object Notation
  - objects and arrays
- Needs to be transformed into a tabular format
- You need to understand the data!
- Data Science Tools:
  - Python/Jupyter/Pandas
  - R

# The New York Times

https://developer.nytimes.com/?it=a

{ } Developers Home APIs Get Started yonen.yonen@yahoo.de

## The New York Times Developer Network

All the **APIs** Fit to Post

GET STARTED

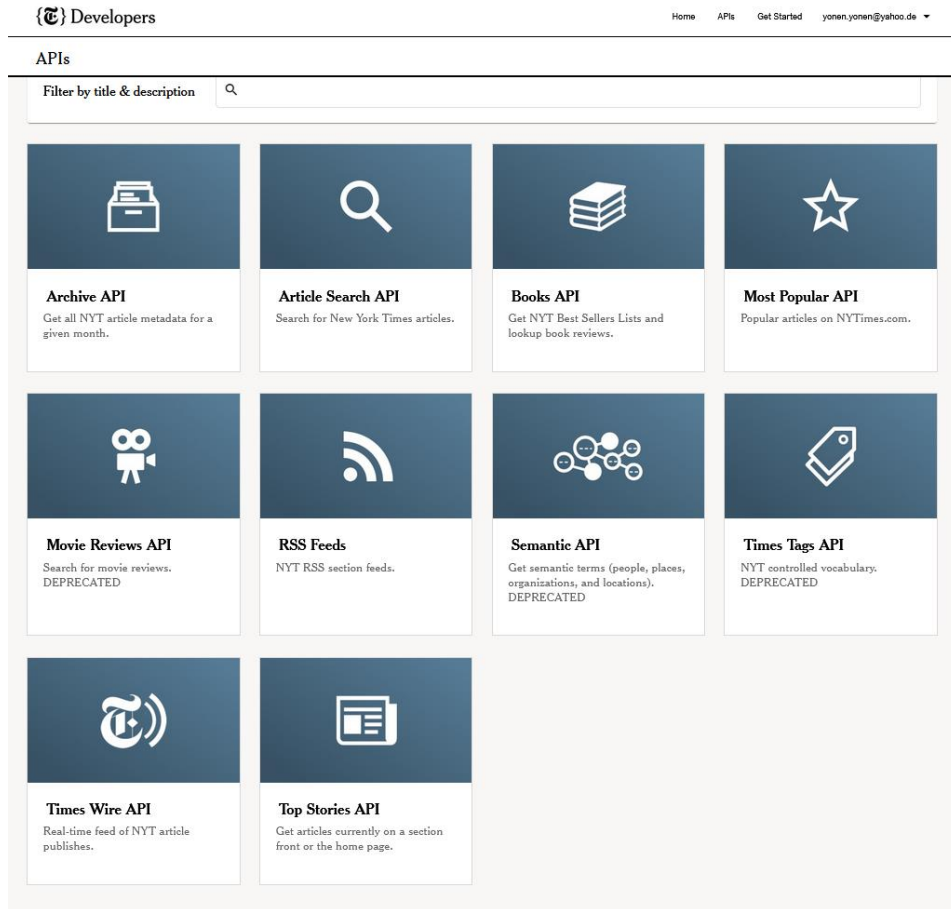
- Get Started**  
Learn how to sign up for an API key.
- APIs**  
Learn about and try out NYT's APIs.
- FAQ**  
Get answers to frequently asked questions.
- Branding**  
Read about NYT's branding & attribution requirements.

Home APIs Terms of Service Branding

- Need to make an account within the Developer Network
- Create an „app“ -> that gives you authorization

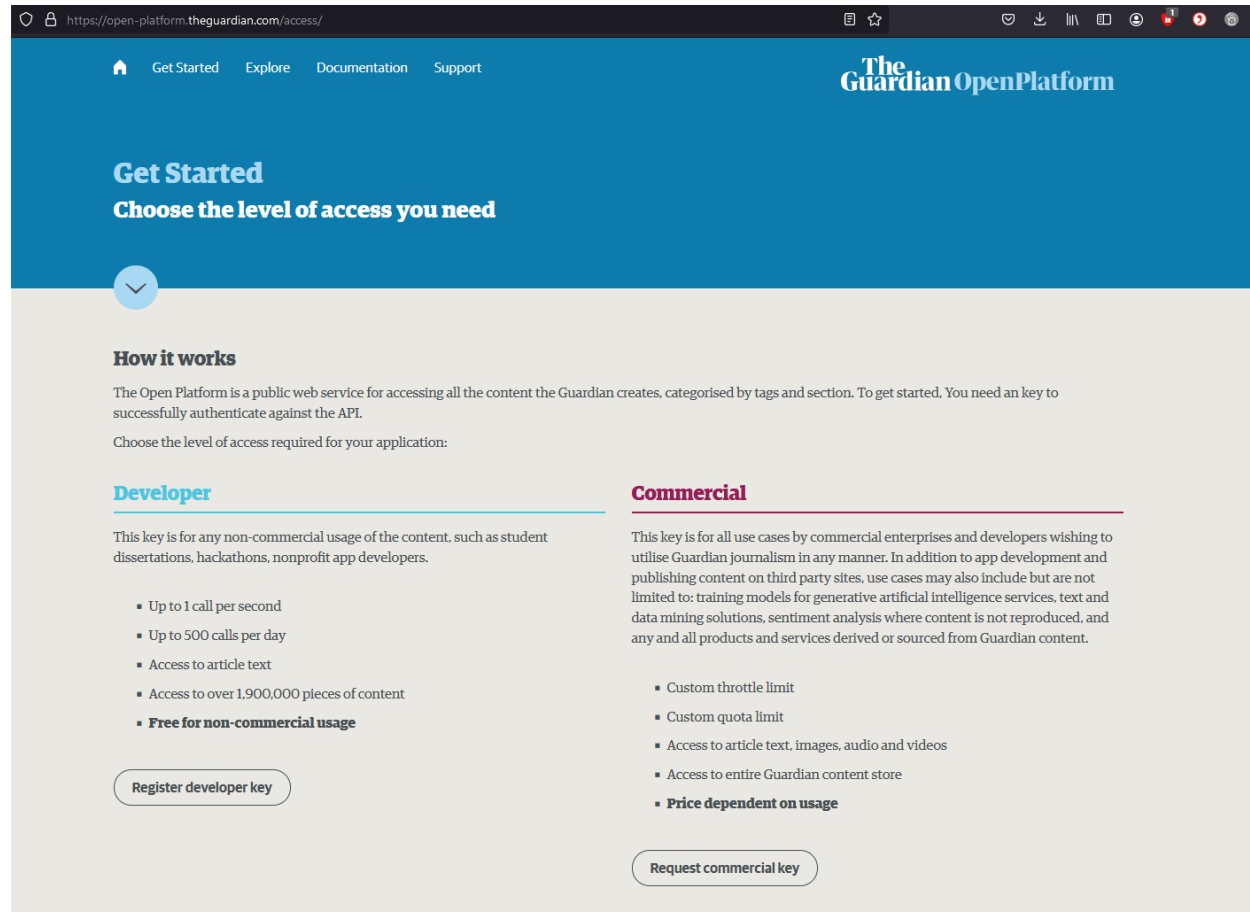


# The New York Times



- Many different APIs
- Archive and Article Search API most interesting for research

# The Guardian



The screenshot shows the 'The Guardian Open Platform' website. The header is blue with navigation links: 'Get Started', 'Explore', 'Documentation', and 'Support'. The main heading is 'Get Started' with the sub-heading 'Choose the level of access you need'. Below this is a section titled 'How it works' which explains that the platform is a public web service for accessing Guardian content, categorized by tags and section. It states that users need a key to authenticate against the API and choose the level of access required for their application. Two options are presented: 'Developer' and 'Commercial'. The 'Developer' option is highlighted with a blue underline and includes a list of features: up to 1 call per second, up to 500 calls per day, access to article text, and access to over 1,900,000 pieces of content. It is noted as 'Free for non-commercial usage' and has a 'Register developer key' button. The 'Commercial' option is highlighted with a red underline and includes a list of features: custom throttle and quota limits, access to article text, images, audio, and videos, access to the entire Guardian content store, and that it is 'Price dependent on usage'. It has a 'Request commercial key' button.

https://open-platform.theguardian.com/access/

Get Started Explore Documentation Support

## The Guardian Open Platform

### Get Started

Choose the level of access you need

#### How it works

The Open Platform is a public web service for accessing all the content the Guardian creates, categorised by tags and section. To get started, You need an key to successfully authenticate against the API.

Choose the level of access required for your application:

#### Developer

This key is for any non-commercial usage of the content, such as student dissertations, hackathons, nonprofit app developers.

- Up to 1 call per second
- Up to 500 calls per day
- Access to article text
- Access to over 1,900,000 pieces of content
- **Free for non-commercial usage**

Register developer key

#### Commercial

This key is for all use cases by commercial enterprises and developers wishing to utilise Guardian journalism in any manner. In addition to app development and publishing content on third party sites, use cases may also include but are not limited to: training models for generative artificial intelligence services, text and data mining solutions, sentiment analysis where content is not reproduced, and any and all products and services derived or sourced from Guardian content.

- Custom throttle limit
- Custom quota limit
- Access to article text, images, audio and videos
- Access to entire Guardian content store
- **Price dependent on usage**

Request commercial key

- Need to register for a key as well

# What do you get access to?

## NYT

- Headlines
- Keywords
- Multimedia
- Abstract

## The Guardian

- Headlines
- Keywords
- Full Text

# What can you do with it?

- Descriptive statistics
- Interrelate with other data

# The World According to Online Newspapers

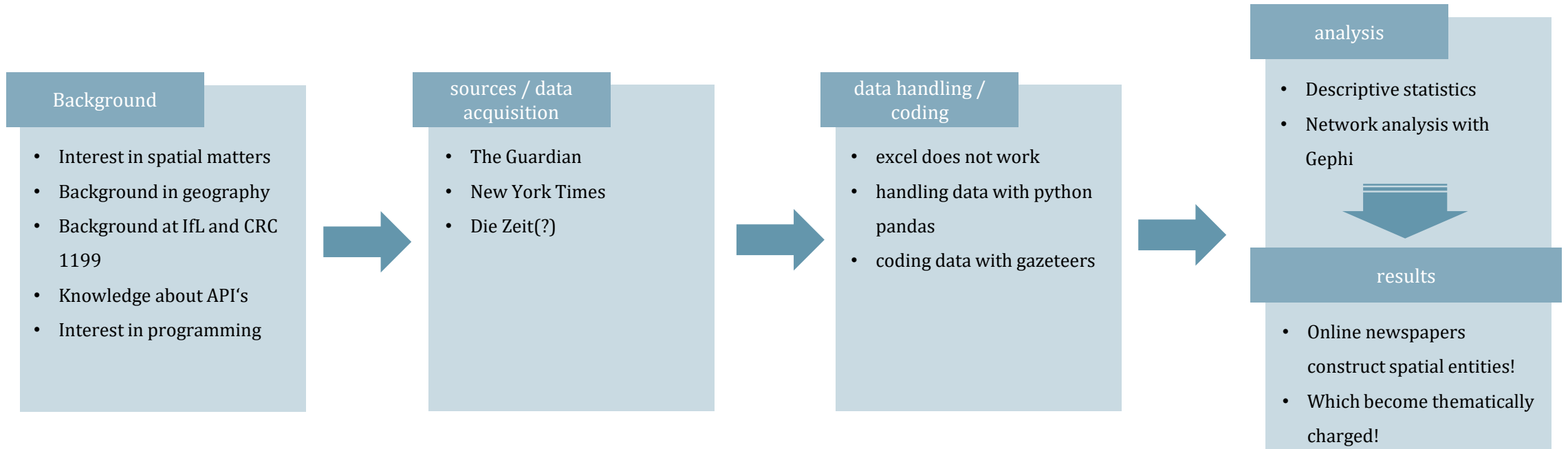




Figure 4 - Guardian Sections in different Editions



Figure 5 - New York Times Editions and Sections

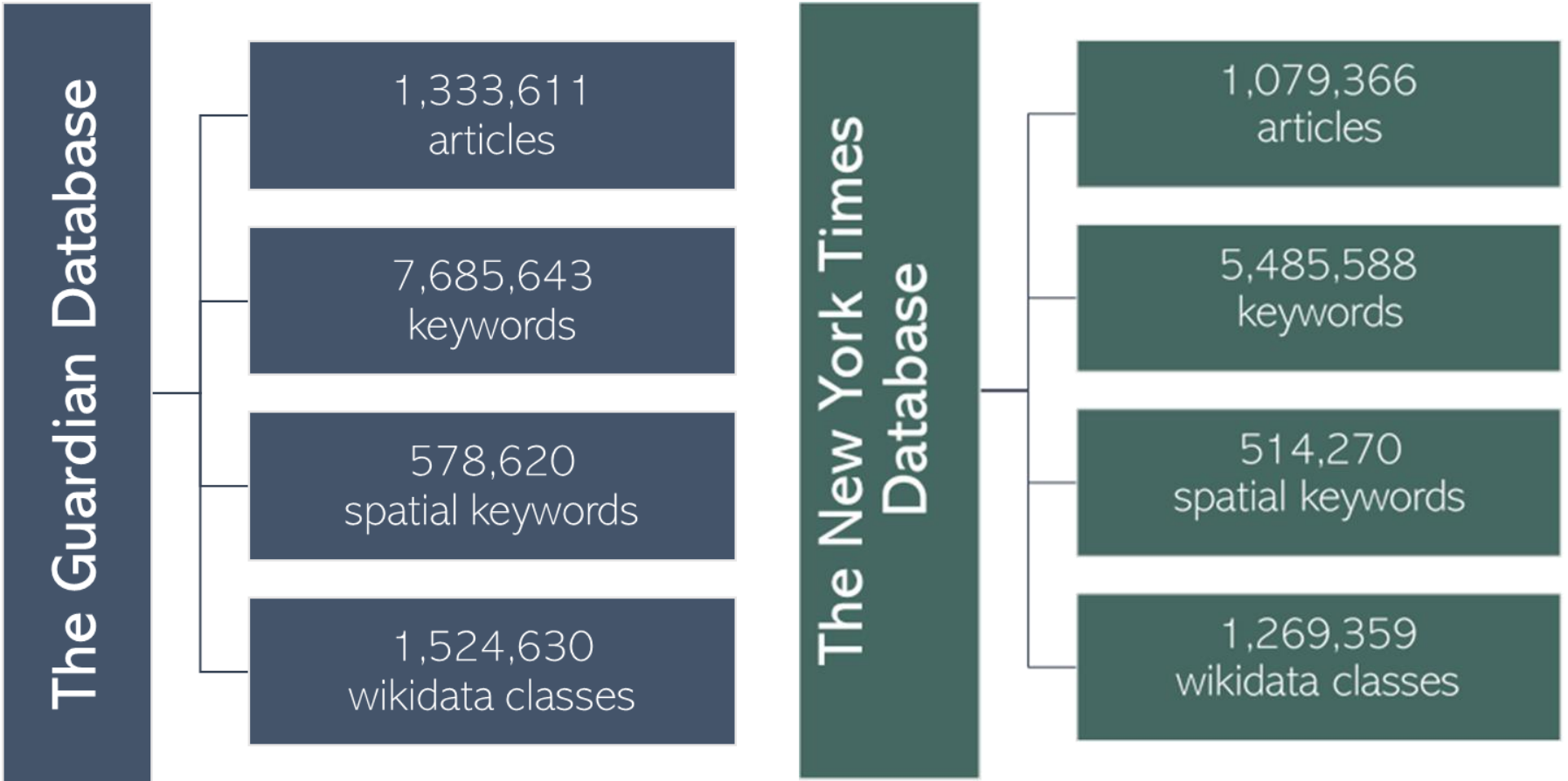


Figure 6 - General Characteristics of the Dataset of The Guardian and the New York Times

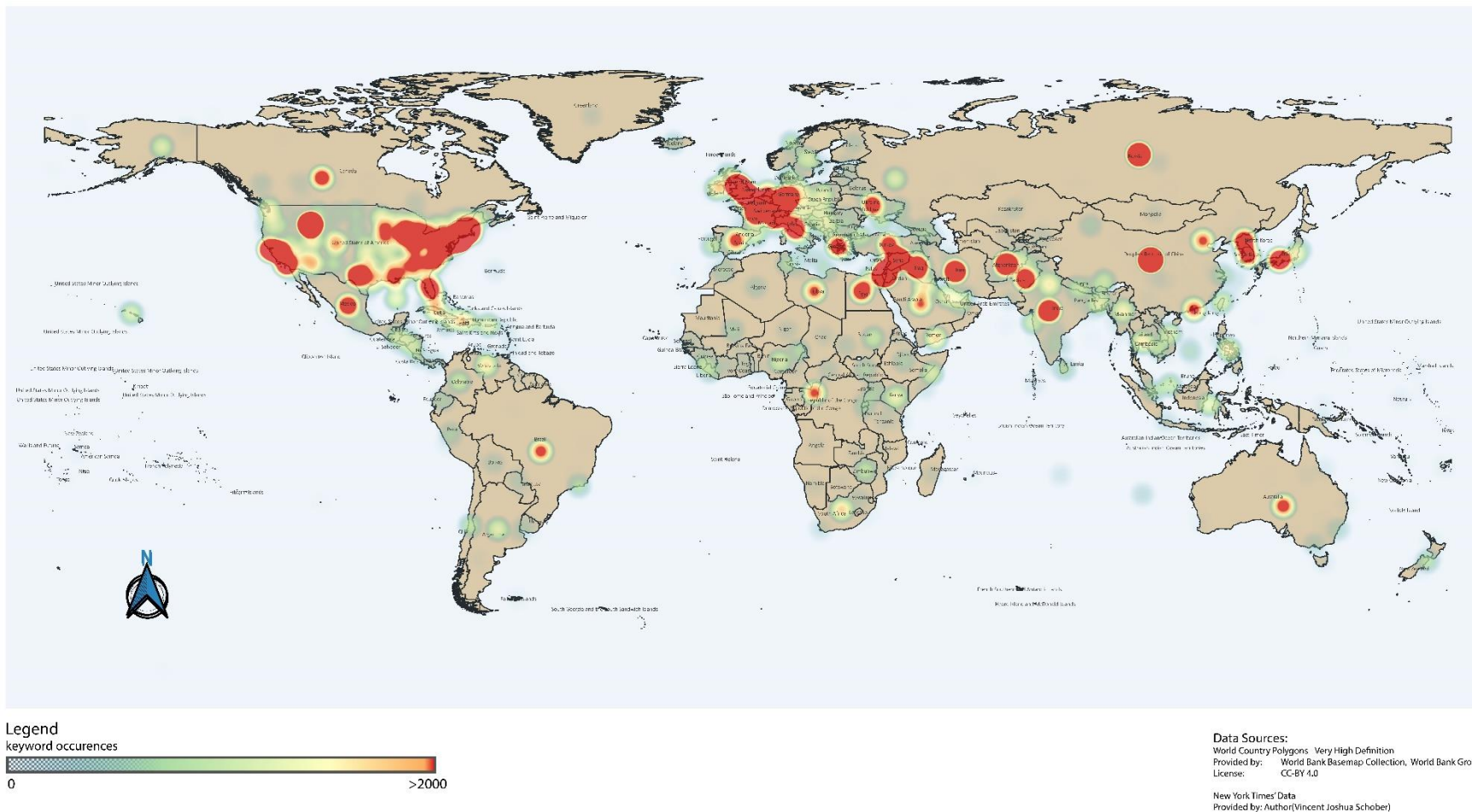
### Heatmap of Spatial Keyword Locations - New York Times

The

S

#### Background

- Interest in s
- Background
- Background
- 1199
- Knowledge
- Interest in p



istics  
is with

pers  
al entities!  
thematically

Figure 31 - New York Times - Heatmap of spatial Keyword Locations

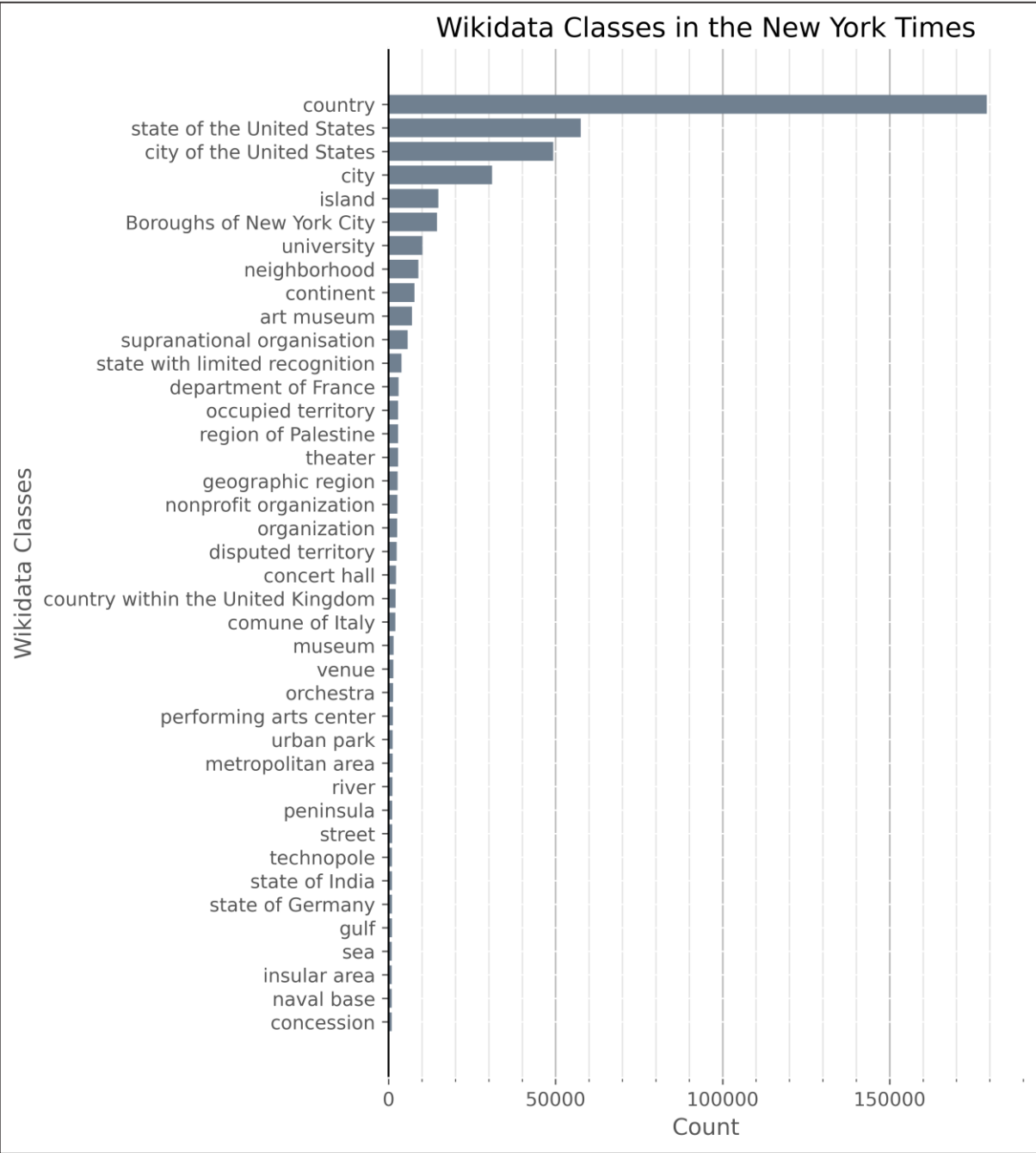


Figure 32 - New York Times - Wikidata Classes



# Other Interesting APIs

- Google Books API
- UNESCO API
- JSTOR Data for Research
- Wikipedia API
- OpenStreetMap API
- Archive.org API

# Thank You!

Vincent J. Schober  
vincent\_j.schober@uni-leipzig.de